REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-
0714

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searc[h] ... iewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including sugges[tions] ... mation Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperv[ork]

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December, 1998

3. REPOR[T]

4. TITLE AND SUBTITLE
1998 Summer Research Program (SRP), Summer Research Extension Program (SREP), Final Report, Volume 1, Armstrong Laboratory and Program Management

5. FUNDING NUMBERS
F49620-93-C-0063

6. AUTHOR(S)
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research & Development Laboratories (RDL)
5800 Uplander Way
Culver City, CA 90230-6608

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research (AFOSR)
801 N. Randolph St.
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The United States Air Force Summer Research Program (SRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members, graduate students, and high school students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates and Air Force Air Logistics Centers (ALC). AFOSR also offers its research associates (faculty only) an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. This volume consists of the SREP program background, management information, statistics, a listing of the participants, and the technical report for each participant of the SREP working at the AF Armstrong Laboratory.

14. SUBJECT TERMS
Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract
G - Grant
PE - Program Element
PR - Project
TA - Task
WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
Leave blank.
NASA - Leave blank.
NTIS -

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self- explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1998

SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 1

PROGRAM MANAGEMENT REPORT

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Colonel Cerveny

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1998

PREFACE

This volume is part of a four-volume set that summarizes the research of participants in the 1998 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 1 of 5, presents the final reports of SREP participants at Armstrong Laboratory.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a 35 page management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1	Armstrong Research Laboratory
2	Phillips Research Laboratory
3	Rome Research Laboratory
4	Wright Research Laboratory
5	Air Logistics Center Arnold Engineering Development Center

1998 SREP Final Technical Report Table of Contents

Armstrong Research Laboratory

Volume 1

1998 SREP Final Technical Report Table of Contents

Phillips Research Laboratory
Volume 2

1998 SREP Final Technical Report Table of Contents

Rome Research Laboratory

Volume 3

1998 SREP Final Technical Report Table of Contents

Wright Research Laboratory

Volume 4

1998 SREP Final Technical Report Table of Contents

Volume 5

1998 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The Summer Research Extension Program (SREP) is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations working at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Appendix 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative charges institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally $18,000 for a $20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1998 SREP PROGRAM

<u>SELECTION DATA</u>: A total of 490 faculty members (SFRP Associates) and 202 graduate students (GSRP associates) applied to participate in the 1998 Summer Research Program. From these applicants 188 SFRPs and 98 GSRPs were selected. The education level of those selected was as follows:

1997 SRP Associates, by Degree			
SFRP		GSRP	
PHD	MS	MS	BS
184	6	2	53

Of the participants in the 1997 Summer Research Program 90 percent of SFRPs and 13 percent of GSRPs submitted proposals for the SREP. One undred and thirty-two proposals from SFRPs and seventeen from GSRPs were selected for funding, which equates to a selection rate of 54% of the SFRP proposals and of 34% for GSRP proposals.

1998 SREP: Proposals Submitted vs. Proposals Selected			
	Summer 1997 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	188	132	20
GSRP	98	17	4
TOTAL	286	149	24

The funding was provided as follows:

Contractual slots funded by AFOSR		18
Laboratory funded		<u>22</u>
	Total	40

Twelve HBCU/MI associates from the 1997 summer program submitted SREP proposals; six were selected (none were lab-funded; all were funded by additional AFOSR funds).

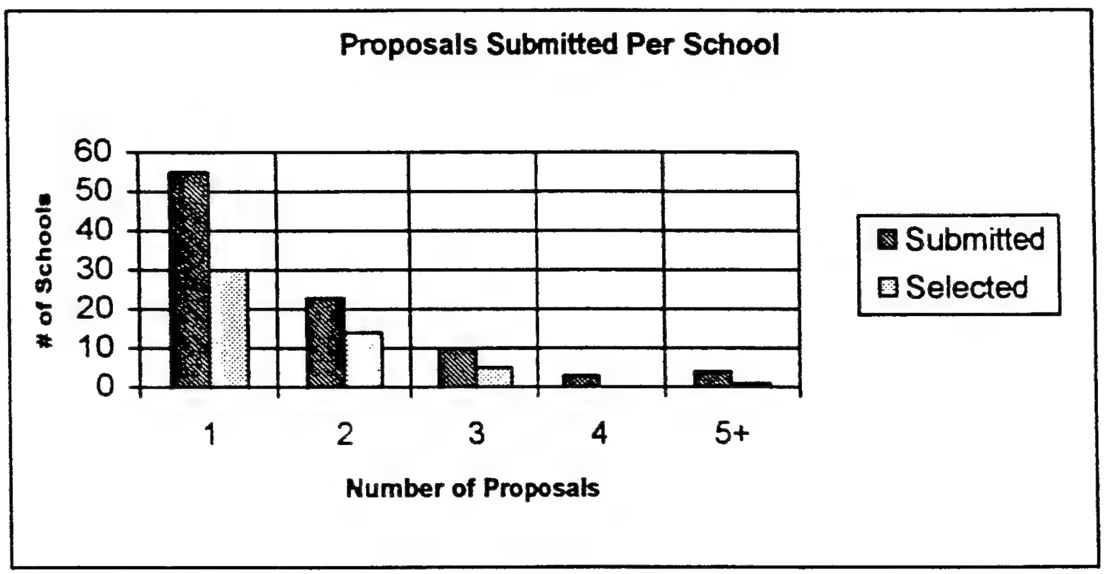
Proposals Submitted and Selected, by Laboratory		
	Applied	Selected
Armstrong Research Site	9	3
Air Logistic Centers	31	5
Arnold Engineering Development Center	2	1
Phillips Research Site	30	10
Rome Research Site	29	12
Wilford Hall Medical Center	1	0
Wright Research Site	47	9
TOTAL	149	40

Note: Armstrong Research Site funded 1 SREP; Phillips Research Site funded 6; Rome Research Site funded 9; Wright Research Site funded 6.

The 125 1997 Summer Research Program participants represented 60 institutions.

Institutions Represented on the 1997 SRP and 1998 SREP		
Number of schools represented in the Summer 97 Program	Number of schools represented in submitted proposals	Number of schools represented in Funded Proposals
125	110	55

Thirty schools had more than one participant submitting proposals.



The selection rate for the 65 schools submitting 1 proposal (68%) was better than those submitting 2 proposals (61%), 3 proposals (50%), 4 proposals (0%) or 5+ proposals (25%). The 4 schools that submitted 5+ proposals accounted for 30 (15%) of the 149 proposals submitted.

Of the 149 proposals submitted, 130 offered institution cost sharing. Of the funded proposals which offered cost sharing, the minimum cost share was $3046.00, the maximum was $39,261.00 with an average cost share of $11,069.21.

Proposals and Institution Cost Sharing		
	Proposals Submitted	Proposals Funded
With cost sharing	117	32
Without cost sharing	32	8
Total	149	40

The SREP participants were residents of 31 different states. Number of states represented at each laboratory were:

States Represented, by Proposals Submitted/Selected per Laboratory		
	Proposals Submitted	Proposals Funded
Armstrong Laboratory	31	5
Air Logistic Centers	9	3
Arnold Engineering Development Center	2	1
Phillips Laboratory	30	10
Rome Laboratory	29	12
Wilford Hall Medical Center	1	0
Wright Laboratory	47	9

Nine of the 1997 SREP Principal Investigators also participated in the 1998 SREP.

<u>ADMINISTRATIVE EVALUATION</u>: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions provided to them by RDL. Thirty-seveb final reports have been received and are included in this report. The subcontracts were funded by $992,855.00 of Air Force money. Institution cost sharing totaled $354.215.00.

TECHNICAL EVALUATION: The form used for the technical evaluation is provided as Appendix 2. Thirty-five evaluation reports were received. Participants by laboratory versus evaluations submitted is shown below:

	Participants	Evaluations	Percent
Armstrong Laboratory	5	4	95.2
Air Logistic Centers	3	3	100
Arnold Engineering Development Center	1	1	100
Phillips Laboratory	10	10	100
Rome Laboratory	12	12	100
Wright Laboratory	9	5	91.9
Total	40	35	95.0

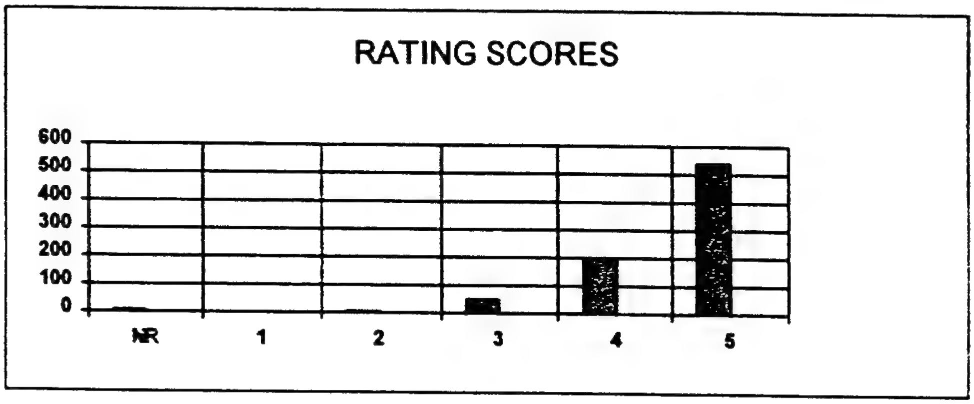
Notes:

1: Research on four of the final reports was incomplete as of press time so there aren't any technical evaluations on them to process, yet. Percent complete is based upon 20/21=95.2%

2: One technical evaluation was not completed because one of the final reports was incomplete as of press time. Percent complete is based upon 18/18=100%

The number of evaluations submitted for the 1998 SREP (95.0%) shows a marked improvement over the 1997 SREP submittals (65%).

PROGRAM EVALUATION: Each laboratory focal point evaluated ten areas (see Appendix 2) with a rating from one (lowest) to five (highest). The distribution of ratings was as follows:

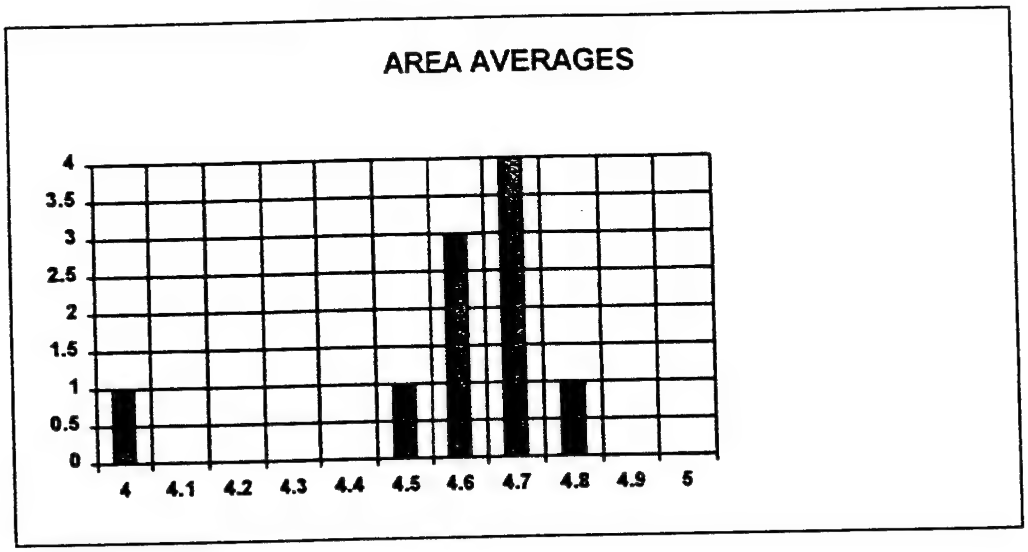


Rating	Not Rated	1	2	3	4	5
# Responses	7	1	7	62 (6%)	226 (25%)	617 (67%)

The 8 low ratings (one 1 and seven 2's) were for question 5 (one 2) "The USAF should continue to pursue the research in this SREP report" and question 10 (one 1 and six 2's) "The one-year period for complete SREP research is about right", in addition over 30% of the threes (20 of 62) were for question ten. The average rating by question was:

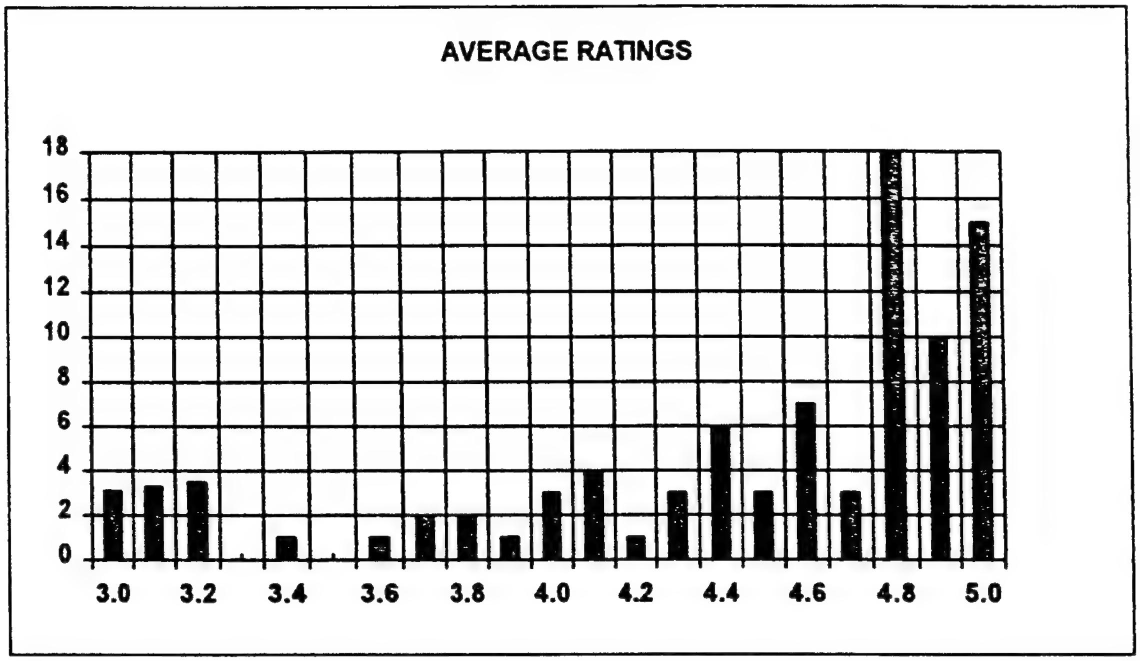
Question	1	2	3	4	5	6	7	8	9	10
Average	4.6	4.6	4.7	4.7	4.6	4.7	4.8	4.5	4.6	4.0

The distribution of the averages was:



Area 10 "the one-year period for complete SREP research is about right" had the lowest average rating (4.1). The overall average across all factors was 4.6 with a small sample standard deviation of 0.2. The average rating for area 10 (4.1) is approximately three sigma lower than the overall average (4.6) indicating that a significant number of the evaluators feel that a period of other than one year should be available for complete SREP research.

The average ratings ranged from 3.4 to 5.0. The overall average for those reports that were evaluated was 4.6. Since the distribution of the ratings is not a normal distribution the average of 4.6 is misleading. In fact over half of the reports received an average rating of 4.8 or higher. The distribution of the average report ratings is as shown:



It is clear from the high ratings that the laboratories place a high value on AFOSR's Summer Research Extension Programs.

3.0 SUBCONTRACTS SUMMARY

Table 1 provides a summary of the SREP subcontracts. The individual reports are published in volumes as shown:

Laboratory	Volume
Armstrong Research Site	1
Arnold Engineering Development Center	5
Air Logistic Centers	5
Phillips Research Site	2
Rome Research Site	3
Wright Research Site	4

SREP SUB-CONTRACT DATA

Report Author / Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Chubb, Gerald Industrial Engineering Ohio State University, Columbus, OH	PhD 98-0829	AL/HR Scoring Pilot Performance of Basic Flight Manuevers	01/01/98 12/31/98	$25000.00	$0.00
Foy, Brent Medical Physics Wright State University, Dayton, OH	PhD 98-0828	AL/OE Development & Validation of a Physiologically-Based Kinetic Model of Perfused	01/01/98 12/31/98	$25000.00	$11278.00
Lance, Charles Psychology Univ of Georgia Res Foundation, Athens, GA	PhD 98-0842	AL/HR Extension of Job Performance Measurement Tech to the Development of a Prototype	01/01/98 12/31/98	$24989.00	$0.00
Woehr, David Department of Psychology Texas A & M Univ-College Station, College	PhD 98-0802	AL/HR Validation of The Multidimensional work ethic profile (MWEP) as a screening too	01/01/98 12/31/98	$25000.00	$11508.00
Collins, Frank Mechanical Engineering Tennessee Univ Space Institute, Tullahoma, TN	PhD 98-0807	AEDC/E Monte Carlo Computation of SpeciesSepaaration by a Conical Skimmer in Hypersonic	01/01/98 12/31/98	$25000.00	$16104.00
Whaley, Paul Mechanical Engineering Oklahoma Christian Univ of Science & Art,	PhD 98-0820	ALC/OC Probabilistic Analysis of Residual Strength in Corroded and Uncorroded Aging Air	01/01/98 12/31/98	$23351.00	$3046.00
Balas, Mark Applied Math Univ of Colorado at Boulder, Boulder, CO	PhD 98-0816	PL/SX Non-Linear Adaptive Control for a Precision Deployable Structure with White ligh	01/01/98 12/31/98	$25000.00	$0.00
Duric, Neb Astrophysics University of New Mexico, Albuquerque, NM	PhD 98-0808	PL/LI Image Recovery Using Phase Diversity	01/01/98 12/31/98	$25000.00	$5777.00
Hanson, George Electrical Engineering Univ of Wisconsin - Milwaukee, Milwaukee, WI	PhD 98-0811	PL/WS Perturbation Analysis of the Natural Frequencies Targets in Inhomogeneous Media	01/01/98 12/31/98	$25000.00	$23250.00
Jeffs, Brian Electrical Engineering Brigham Young University, Provo, UT	PhD 98-0813	PL/LI Bayesian Restoration of Space object Images From Adaptive Optics Data with unkno	01/01/98 12/31/98	$25000.00	$19177.00
Kar, Aravinda Engineering University of Central Florida, Orlando, FL	PhD 98-0812	PL/LI Effects of Vapor-Plasma Layer on Thick-Section Cutting and Calculation of Modes	01/01/98 12/31/98	$25000.00	$5414.00
Leo, Donald Mechanical & Aerospace University of Toledo, Toledo, OH	PhD 98-0810	PL/VT Adaptive vibration suppression for autonomous Control Systems	01/01/98 09/30/98	$24964.00	$9628.00
Liu, Hanli Physics Univ of Texas at Arlington, Arlington, TX	PhD 98-0814	PL/LI Continuous-Wave Approach to 3-D Imaging through Turbid media w/a Single Planar M	01/01/98 12/31/98	$25000.00	$11000.00
Bienfang, Joshua Physics University of New Mexico, Albuquerque, NM	BS 98-0815	PL/LI Optical Clocks Based on Diode Lasers	01/01/98 12/31/98	$24994.00	$0.00
Paulson, Eric Engineering/Physics Univ of Colorado at Boulder, Boulder, CO	BS 98-0837	PL/RK Optimization & Analysis of a Waverider Vehicle for Global Spaceplane Trajectorie	01/01/98 12/31/98	$25000.00	$7794.00

SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount	Univ. Cost Share
Stephens II , Kenneth University of North Texas, Denton, TX	MA 98-0809	PL/WS Simulation of an Explosively Formed Fuse Using MACH 2	01/01/98	12/31/98	$25000.00	$16764.00
Barjaktarovic , Milica Electrical Engineering Wilkes University, Wilkes Barre, PA	PhD 98-0824	RL/IW Specification and Verification of SDN.701 MSP Functions and Missi Crypto Functio	01/01/98	12/31/98	$24976.00	$3158.00
Batalama , Stella EE SUNY Buffalo, Buffalo, NY	PhD 98-0823	RL/C3 Robust Spread Spectrum Communications:Adaptive Interference Mitigation Technique	01/01/98	12/31/98	$25000.00	$5600.00
Bourbakis , Nikolaos Computer Science & Engr SUNY Binghamton, Binghamton, NY	PhD 98-0832	RL/IR hierarchical-Adaptive Image Segmentation	01/01/98	12/31/98	$25000.00	$22723.00
Dasigi , Venugopala Computer Science Southern Polytechnic State Univ, Marietta, GA	PhD 98-0830	RL/C3 Information Fusion w/Multiple Feature Extractors for automatic Text Classificati	01/01/98	12/31/98	$25000.00	$4000.00
Eckert , Richard Physics SUNY Binghamton, Binghamton, NY	PhD 98-0825	RL/C3 The Interactive Learning Wall; A PC-Based, Deployable Data Wall for Use in a Co	01/01/98	12/31/98	$25000.00	$39261.00
Lin , Kuo-Chi Aerospace Engineering University of Central Florida, Orlando, FL	PhD 98-0822	RL/IR Web-Based Distributed Simulation	01/01/98	12/31/98	$25000.00	$0.00
Pados , Dimitrios Dept. of Electrical /Computer Eng. State Univ. of New York Buffalo, Buffalo, NY	PhD 98-0818	RL/OC Adaptive Array Radars and Joint Space-Time Auxiliary Verctor Filtering	01/01/98	12/31/98	$25000.00	$5600.00
Panda , Brajendra Computer Science University of North Dakota, Grand Forks, ND	PhD 98-0821	RL/CA Information Warfare: Design of an Efficient Log Management Method to Aid In Dat	01/01/98	12/31/98	$25000.00	$7113.00
Pittarelli , Michael Systems Science SUNY OF Tech Utica, Utica, NY	PhD 98-0827	RL/C3 Complexity of Detecting and content-driven methods for resolving database incons	01/01/98	12/31/98	$24998.00	$0.00
Schmalz , Mark Dept of Computer & Info Science University of Florida, Gainesville, FL	PhD 98-0831	RL/IR Errors Inherent in 3D Target Reconstruction from Multiple Airborne Images	01/01/98	12/31/98	$24619.00	$0.00
Ye , Nong Industrial Engineering Arizona State University, Tempe, AZ	PhD 98-0826	RL/CA Model-Based Assessment of Campaign Plan Performance under Uncertainty	01/01/98	12/31/98	$25000.00	$5000.00
Bradley , Parker Physics Syracuse University, Syracuse, NY	BS 98-0834	RL/IR Development of User-Friendly Comp Environment for Blind Source Separation Studie	01/01/98	12/31/98	$25000.00	$0.00
Kumar , Devendra Computer Science CUNY-City College, New York, NY	PhD 98-0805	ALC/SA Further Development of a Simpler, Multiversion concurrency Control Protocol for	01/01/98	12/31/98	$25000.00	$11362.00
Chow , Joe Mechanical Engineering Florida International Univ, Miami, FL	PhD 98-0806	ALC/W An Automated 3-D Surface Model Creation Module for Laser Scanned Point Data	01/01/98	12/31/98	$25000.00	$5360.00

SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Beecken , Brian Physics Bethel College, St. Paul, MN	PhD 98-0804	WL/MN	01/01/98 12/31/98	$19986.00	$3997.00
		Development of a statistical Model predicting the impact of a scene projector's			
Beggs , John Electrical Engineering Mississippi State University, Mississippi State,	PhD 98-0817	WL/FI	01/01/98 12/31/98	$25000.00	$25174.00
		Implementation of an Optimization Algorithm in Electromagnentics for Radar Absor			
Bhatnagar , Raj Computer Science University of Cincinnati, Cincinnati, OH	PhD 98-0819	WL/AA	01/01/98 09/30/98	$25000.00	$17488.00
		Analysis of Intra-Class Variability & synthetic Target Models for Use in ATR			
Blaisdell , Gregory Mechanical Engineering Purdue University, West Lafayette, IN	PhD 98-0839	WL/FI	01/01/98 12/31/98	$25000.00	$11844.00
		Validation of a Large Eddy Simulation Code & Development of Commuting Filters			
Douglass , John Zoology University of Arizona, Tucson, AZ	PhD 98-0803	WL/MN	01/01/98 12/31/98	$25000.00	$3719.00
		Roles of Matched Filtering and Coarse in Insect Visual Processing			
Hosford , William Mettalurgy Univ of Michigan, Ann Arbor, MI	PHD 98-0840	WL/MN	01/01/98 12/31/98	$25000.00	$5000.00
		Prediction of Compression Textures in Tantalum Using a Pencil-Glide Computer Mod			
Pan , Yi Computer Science University of Dayton, Dayton, OH	PhD 98-0838	WL/FI	01/01/98 12/31/98	$25000.00	$9486.00
		Parallelization of Time-Dependent Maxwell Equations Using High Perform. Fortran			
Pochiraju , Kishore Mechanical Engineering Stevens Inst of Technology, Hoboken, NJ	PhD 98-0833	WL/ML	01/01/98 12/31/98	$25000.00	$9625.00
		A Hybrid Variational-Asymptotic Method for the Analysis of MicroMechanical Damag			
Shtessel , Yuri Electrical Engineering Univ of Alabama at Huntsville, Huntsville, AL	PhD 98-0841	WL/FI	01/01/98 12/31/98	$25000.00	$4969.00
		Continuous Sliding Mode Control Approach for Addressing actutor Deflection and			
Starzyk , Janusz Electrical Engineering Ohio University, Athens, OH	PhD 98-0801	WL/AA	01/01/98 12/31/98	$24978.00	$12996.00
		Feature Selection for Automatic Target Recognition:Mutual Info & Stat Tech			

APPENDIX 1:

SAMPLE SREP SUBCONTRACT

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1998 SUMMER RESEARCH EXTENSION PROGRAM
SUBCONTRACT 98-0812

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

University of Central Florida
Office of Sponsored Research/ Admin#423
4000 Central Florida Blvd.
Orlando, FL 32816-0150

REFERENCE: Summer Research Extension Program Proposal 97-0018

Start Date: 01/01/98 End Date 12/31/98

Proposal Amount: $25000.0
Proposal Title:
Effects of Vapor-Plasma Layer on Thick-Section Cutting and Calculation of Modes

(1) PRINCIPAL INVESTIGATOR:
DR Aravinda Kar
CREOL
University of Central Florida
Orlando, FL 32816-2700

(2) UNITED STATES AFOSR CONTRACT NUMBER: F49620-93-C-0063

(3) CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA): 12.800
PROJECT TITLE: AIR FORCE DEFENCE RESEARCH SOURCES PROGRAM

(4) ATTACHMENTS
1 REPORT OF INVENTIONS AND SUBCONTRACT
2 CONTRACT CLAUSES
3 FINAL REPORT INSTRUCTIONS

***** SIGN SREP SUBCONTRACT AND RETURN TO RDL *****

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-93-C-0063) to the United States Air Force to administer the Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Summer Research Extension Program (SREP). Approximately 61 SREP proposals annually will be selected by the Air Force for funding of up to $25,000; shared funding by the academic institution is encouraged. SREP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents an agreement between RDL and the institution herein designated in Section 5 below.

2. RDL PAYMENTS: RDL will provide the following payments to SREP institutions:

- 80 percent of the negotiated SREP dollar amount at the start of the SREP research period.
- The remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the SREP research.

3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

a. Assure that the research performed and the resources utilized adhere to those defined in the SREP proposal.

b. Provide the level and amounts of institutional support specified in the SREP proposal..

c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.

d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract, but no later than December 31, 1998. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1998.

e. Assure that the final report is submitted in accordance with Attachment 3.

f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade associations meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.

g. Notify RDL of inventions or patents claimed as the result of this research as specified in Attachment 1.

h. RDL is required by the prime contract to flow down patent rights and technical data requirements to this subcontract. Attachment 2 to this subcontract

contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

RDL AFOSR Program Office
5800 Uplander Way
Culver City, CA 90230-6609

5. By their signatures below, the parties agree to provisions of this subcontract.

Abe Sopher RDL Contracts Manager	Signature of Institution Contracting Official
	Typed Printed Name
Date	Title
	Institution
	Date Phone

52.215-12	SUBCONTRACTOR COST OR PRICING DATA
52.215-14	INTEGRITY OF UNIT PRICES
52.215-8	ORDER OF PRECEDENCE
52.215.18	REVERSION OR ADJUSTMENT OF PLANS FOR POSTRETIREMENT BENEFITS OTHER THAN PENSIONS
52.222-3	CONVICT LABOR
52.222-26	EQUAL OPPORTUNITY
52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS
52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERAN AND VETERANS OF THE VIETNAM ERA
52.223-2	CLEAN AIR AND WATER
52.223-6	DRUG-FREE WORKPLACE
52.224-1	PRIVACY ACT NOTIFICATION
52.224-2	PRIVACY ACT
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS
52.227-1	ALT. I - AUTHORIZATION AND CONSENT
52.227-2	NOTICE AND ASSISTANCE REGARDING PATIENT AND COPYRIGHT INFRINGEMENT

52.227-10	FILING OF PATENT APPLICATIONS - CLASSIFIED SUBJECT MATTER
52.227-11	PATENT RIGHTS - RETENTION BY THE CONTRACTOR (SHORT FORM)
52.228-7	INSURANCE - LIABILITY TO THIRD PERSONS
52.230-5	COST ACCOUNTING STANDARDS - EDUCATIONAL INSTRUCTIONS
52.232-23	ALT. I - ASSIGNMENT OF CLAIMS
52.233-1	DISPUTES
52.233-3	ALT. I - PROTEST AFTER AWARD
52.237-3	CONTINUITY OF SERVICES
52.246-25	LIMITATION OF LIABILITY - SERVICES
52.247-63	PREFERENCE FOR U.S. - FLAG AIR CARRIERS
52.249-5	TERMINATION FOR CONVENIENCE OF THE GOVERNMENT (EDUCATIONAL AND OTHER NONPROFIT INSTITUTIONS)
52.249-14	EXCUSABLE DELAYS
52.251-1	GOVERNMENT SUPPLY SOURCES

DOD FAR CLAUSES	DESCRIPTION
252.203-7001	SPECIAL PROHIBITION ON EMPLOYMENT
252.215-7000	PRICING ADJUSTMENTS
252.233-7004	DRUG FREE WORKPLACE (APPLIES TO SUBCONTRACTS WHERE THERE IS ACCESS TO CLASSIFIED INFORMATION)
252.225-7001	BUY AMERICAN ACT AND BALANCE OF PAYMENTS PROGRAM
252.225-7002	QUALIFYING COUNTRY SOURCES AS SUBCONTRACTS
252.227-7013	RIGHTS IN TECHNICAL DATA - NONCOMMERCIAL ITEMS
252.227-7030	TECHNICAL DATA - WITHOLDING PAYMENT
252.227-7037	VALIDATION OF RESTRICTIVE MARKINGS ON TECHNICAL DATA
252.231-7000	SUPPLEMENTAL COST PRINCIPLES
252.232-7006	REDUCTIONS OR SUSPENSION OF CONTRACT PAYMENTS UPON FINDING OF FRAUD

APPENDIX 2:

SAMPLE TECHNICAL EVALUATION FORM

SUMMER RESEARCH EXTENSION PROGRAM
TECHNICAL EVALUATION

SREP No: 98-0810
Principal Investigator: DR Donald Leo
University of Toledo

Circle the rating level number, 1 (low) through 5 (high), you feel best evaluate each statement and return the completed form to RDL by fax of mail to:

RDL
Attn: SREP Tech Evals
5800 Uplander Way
Culver City, CA 90230-6608
(310)216-5940 or (800)677-1363

1.	This SREP report has a high level of technical merit.	1 2 3 4 5
2.	The SREP program is important to accomplishing the lab's mission.	1 2 3 4 5
3.	This SREP report accomplished what the associate's proposal promised.	1 2 3 4 5
4.	This SREP report addresses area(s) important to the USAF.	1 2 3 4 5
5.	The USAF should continue to pursue the research in this SREP report.	1 2 3 4 5
6.	The USAF should maintain research relationships with this SREP associate.	1 2 3 4 5
7.	The money spent on this SREP effort was well worth it.	1 2 3 4 5
8.	This SREP report is well organized and well written.	1 2 3 4 5
9.	I'll be eager to be a focal point for summer and SREP associates in the future.	1 2 3 4 5
10.	The one-year period for complete SREP research is about right.	1 2 3 4 5

11. If you could change any one thing about the SREP program, what would you change:

____

12. What do you definitely NOT change about the SREP program?

____

PLEASE USE THE BACK FOR ANY OTHER COMMENTS

Laboratory Phillips Laboratory
Lab Focal Point Capt Jeanne Sullivan
Office Symbol AFRL/VSDV   Phone: (505)846-2069

SUMMER RESEARCH EXTENSION PROGRAM
TECHNICAL EVALUATION

SREP No: 98-0810
Principal Investigator: DR Donald Leo
University of Toledo

Circle the rating level number, 1 (low) through 5 (high), you feel best evaluate each statement and return the completed form to RDL by fax of mail to:

RDL
Attn: SREP Tech Evals
5800 Uplander Way
Culver City, CA 90230-6608
(310)216-5940 or (800)677-1363

1.	This SREP report has a high level of technical merit.	1 2 3 4 5
2.	The SREP program is important to accomplishing the lab's mission.	1 2 3 4 5
3.	This SREP report accomplished what the associate's proposal promised.	1 2 3 4 5
4.	This SREP report addresses area(s) important to the USAF.	1 2 3 4 5
5.	The USAF should continue to pursue the research in this SREP report.	1 2 3 4 5
6.	The USAF should maintain research relationships with this SREP associate.	1 2 3 4 5
7.	The money spent on this SREP effort was well worth it.	1 2 3 4 5
8.	This SREP report is well organized and well written.	1 2 3 4 5
9.	I'll be eager to be a focal point for summer and SREP associates in the future.	1 2 3 4 5
10.	The one-year period for complete SREP research is about right.	1 2 3 4 5

11. If you could change any one thing about the SREP program, what would you change:

12. What do you definitely NOT change about the SREP program?

PLEASE USE THE BACK FOR ANY OTHER COMMENTS

Laboratory: Phillips Laboratory
Lab Focal Point: Capt Jeanne Sullivan
Office Symbol: AFRL/VSDV Phone: (505)846-2069

SCORING PILOT PERFORMANCE OF BASIC FLIGHT MANEUVERS

Gerald P. Chubb
Associate Professor
Department of Aerospace Engineering and Aviation

The Ohio State University
164 W. 19th Avenue
Columbus, HO 43210

Final Report for:
Summer Research Extension Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force base, DC

and

Armstrong Laboratory
Warfighter Training Division

June 1999

SCORING PILOT PERFORMANCE
OF BASIC FLIGHT MANEUVERS

Gerald P. Chubb
Associate Professor
Department of Aerospace Engineering and Aviation
The Ohio State University

Abstract

The scoring of student pilot performance has typically been done by subjective assessments performed by the student's flight instructor. Many of the maneuvers that need to be learned early in flight training are well-defined. The criteria for acceptable maneuver performance have, at least in some cases, been defined by the Federal Aviation Adminstration (FAA) in their Practical Test Standards (PTS). The purpose of the present study was to use a Commercial Off the Shelf (COTS) flight simulation software package as the target system for developing a quantitative scoring system for evaluating the performance of these basic flight maneuvers.

A windows-based scoring system was developed and demonstrated that allows a student to perform commanded maneuvers and get scored on how well they perform those maneuvers. The scoring criteria and weights placed on individual measures can be set by the user. This will permit further research on how to best set these weights and combine the measures into metrics most meaningful and useful to the student and instructor.

Only limited use has been made of the scoring system, and its utility needs to be tested with a set of actual students and instructors in order to determine how well it is accepted and whether it provides any benefits over the conventional subjective assessment methods now used.

The chosen flight simulation software was Microsoft's Flight Simulator 98 (MS-FS), using a yoke and set of rudder pedals connected to a Personal Computer (PC) through a game card. The windows-based scoring system is designed as a stand alone, third-party software add-on that can be used in conjunction with MS-FS to give the student quantitative scores that reflect the quality of the student's execution of the requested maneuver.

SCORING PILOT PERFORMANCE
OF BASIC FLIGHT MANEUVERS

Gerald P. Chubb

Introduction

To develop a skill and maintain proficiency, the performer needs feedback, indicating how well they performed. Student pilots are the performers of interest in this case. Initially, flight instructors provide knowledge of results by performing a subjective assessment of the student pilot's performance and communicate their assessment to the student verbally. Along the way, the student learns the cues and internalizes the criteria for what constitutes acceptable, if not superior performance.

A better approach is to actually measure pilot performance against some desired flight path, determine the deviations from the ideal, and show the pilot what they did and when they did it. This approach is often used in training private pilots how to perform a precision approach using an Instrument Landing System (ILS), as they progress on to getting their instrument rating. However, comparable scoring is not available for the basic maneuvers a private pilot must learn before being approved for the first solo flight. This is the topic of concern in this study: scoring those basic flight maneuvers. While such scoring is now feasible technically, the question is whether it is useful, and if so, how to do it well: in a fashion acceptable to students and their instructors.

If scoring of this sort can be done with a set of simple maneuvers, then in principle it should be possible to do it with the more complicated maneuvers encountered in aerobatics and in the Basic Flight Maneuvers (BFMs) required to teach combat tactics. Before looking at the BFMs necessary for training Air Force pilots in air-to-air combat tactics, it seemed prudent to develop methods that would apply to measuring a private pilot's performance. The Federal Aviation Administration (FAA) requires that all private pilots be able to perform acceptably well the maneuvers specified in the FAA's Practical Test Standards (PTS). The PTS maneuvers include all of the basic flight maneuvers one must learn to perform in order to control the flight of an aircraft.

The more complicated BFMs that are prerequisite for combat tactics training rest on the pilot's ability to perform the very basic maneuvers identified in the PTS. The military BFMs are more closely related to aerobatic maneuvers, which, if learned at all, come after gaining one's private pilot certificate. While the present work focuses on measuring PTS performance, an obvious extension would be to address the aerobatic manuevers and the military BFMs as the next step.

Instrumenting an aircraft to get quantitative data is an expensive proposition. However, desktop personal computers (PCs) now host a wide variety of flight simulation software. The most popular commercial off the shelf airplane simulation package is Microsoft (MS) Flight Simulator (FS). MS-FS has a broad acceptance among "want-to-be"

pilots, more so than with actual pilots. Actual pilots recognized there were a number of deficiencies with the early versions of MS-FS and in many cases bought something better. Several options are available on the market today.

Because of its popularity, MS-FS has attracted a number of others to develop compatible software, either as third party vendors or as shareware. One such shareware package provides users with the ability to capture data from MS-FS as it is operating. This software provides information about the aircraft's altitude, airspeed, and heading – all of which are important parameters to control in flying an airplane.

Based on this software's ability to monitor MS-FS performance during real time simulations and create a data file of the airplane's performance, it appeared feasible to take this data and construct a scoring system that evaluated how well an individual performed particular flight tasks. By having a numeric score to reflect how well a pilot performed a particular maneuver, both the instructor and the student pilot benefit. Each would have a sound basis for determining whether the student's performance was improving and whether it had improved enough to meet FAA PTS: a requirement if one is to get a private pilot's certificate.

Also, it seems reasonable to hope that having a continuous numeric score on performance, the instructor could also determine where a student was experiencing difficulty learning a maneuver or developing proficiency. This might prove useful in devising an appropriate remediation of the skill deficiency. It is not self-apparent how to construct such aids for the instructor, but having the scoring system in-place is pre-requisite to beginning this kind of research activity.

Moreover, if the scoring system can be applied to a data stream from MS-FS, then in theory it could be equally well applied to a data stream from any other flight simulation software. OSU has two Aviation Simulation Trainers (AST) flight training devices, both a single and a multi-engine model. Since these use a form of Basic as the native programming language, we have already been able to capture and record data from these devices. It therefore appears possible to use the scoring system in our own flight education system, once we prove its utility and validity.

To begin with, we make a distinction between pilot behavior and (aircraft) performance. Behavior is what the pilot does with the control inputs. These include yoke inputs (rotation and longitudinal push / pull of the yoke), rudder pedal inputs, and flap settings. It was assumed the airplane had fixed gear, otherwise gear extension would be another important control. Aircraft performance was the response of the vehicle to these inputs. The pilot's job is to behave in such a fashion that performance is acceptable (safe).

While the real goal is to train pilot behavior in order to assure appropriate aircraft performance, it is still necessary to measure aircraft maneuver performance in order to assure the behavior satisfied a particular goal. However, as the performance scoring concepts were developed, it was recognized that there may also be a need to capture in a more quantitative fashion the behavior of the pilot: what controls were moved to which position at what time.

For example, there is more than one control strategy that can achieve the same aircraft performance. Also, large control inputs at low speeds may not result in large aircraft excursions from optimum but are considered poor pilot flying technique. Small, smooth control inputs are preferred, which often result in only minor changes in aircraft performance. Therefore, a number of subtle questions began to emerge as we examined the scoring issues more carefully.

The behavioral aspects of measuring behavior have not been ignored in this effort, but they were intentionally deferred until it can be shown that the performance measures work. Also, the software that captures aircraft performance does not capture pilot control inputs, unfortunately. Additional development of the data capture software would be required, something that would go beyond the scope of the present study.

Discussion of the Problem

The FAA publishes FAA-S-8081-14, PRIVATE PILOT PRACTICAL TEST STANDARDS (PTS) FOR AIRPLANE, the most recent edition of which is dated May 1995. FAA inspectors and Designated Examiners use this document as the basis for their check ride. The check ride is the last test a student pilot takes before being given the private pilot's certificate (passing prescribed written and oral tests are a required prerequisite to getting this check ride). This FAA PTS document provides one basis for establishing objective maneuver criteria. However, the rationale for those criteria is not provided. Therefore some of the criteria appear on the surface to be arbitrary. We will attempt to rationalize at least some of the criteria.

Other publicly available government documents also proved useful in developing our scoring system. Some of the parameters included in the scoring system are taken from several of the FAA flight training publications such as AC 61-21, FLIGHT TRAINING HANDBOOK, and AC 61-23, PILOT'S HANDBOOK OF AERONAUTICAL KNOWLEDGE and the AIM (Airman's Information Manual.) Reference was also made to commercially available publications. Col. (Ret.)Vogel's personal flight instructing experience also played a role in formulating selected aspects of the scoring system.

The scoring parameters are applied to four basic flight training maneuvers that underlie all of the conditions of flight that can be encountered. By training each student pilot to flawlessly accomplish each of these basic maneuvers, the flight instructor can then help the trainee combine them into more complex maneuvers that one needs to know to become an accomplished and safe pilot.

The four basic maneuvers are: 1) climbs, 2) descents, 3) turns, and 4) straight and level flight. Straight and level flight is described as a series of very small climbs, descents, and turns to maintain a line through the sky. More complex maneuvers such as climbing and descending turns, constant rate maneuvers, constant speed maneuvers and

other complex tasks are simply combinations of the four basic maneuvers. How well the student performs depends upon how well a set of specified parameters is controlled.

This rest of this section discusses some of the alternative scoring methods that have been used. The methods discussed here are quantitative methods applied typically to laboratory data on various manual tracking tasks. Two type of such tasks have typically been used: compensatory and pursuit tracking. In compensatory tracking, the subject's task is to null an error indication.

Compensatory tracking is an analogue for several different kinds of flying tasks. The most common in civil aviation is the precision approach task using an ILS. However, there are several important differences. First, the real task is a two-axis task with cross-coupled dynamics: changes in one axis can and do affect dynamics in the other. Second, in the real world, the forcing function is usually wind speed and direction, which may be constant, variable, or gusting. Most laboratory tasks have used simpler dynamics and a random disturbance function to emulate the impact of uncertainties in wind speed and direction. Compensatory tracking is also an analogue of a strategic bomb run: the navigator's crosshairs are centered on an aiming point, and the heading error indicator tells the pilot what must be done to drive over the release point.

In pursuit tacking, a cursor is to be place over a moving indicator. This is an analogue of an air-to-air engagement, where the pilot is chasing another aircraft and must get within the gun or missile envelope. Although pursuit tracking appears to be the more difficult of these two tasks, studies have typically indicated that pursuit tracking is done with less error than compensatory tracking.

In many of the early tracking studies, the task was implemented on an analogue computer. The raw data was an electronic signal (e.g. voltage) that was measured. This influenced, to some degree, the kind of data collected and the scoring of that data. The signal could typically be scaled to represent whatever variable was of interest (pitch, bank, altitude, airspeed, angle-off, etc.).

By using a strip chart recorder, the voltage level could be used to move a pen on a roll of moving paper, such that a time trajectory was generated that graphically showed how the data measure changed (increased / decreased) over time. The pen trace provided a visible record of measured values over time. This state trajectory provided a detailed description of performance, but it was not easy to analyze numerically. The pen trace had to be converted to digital data in order to subject the measures to analysis. Doing that was labor intensive: it took a lot of time!

To escape this labor intensive analysis, other methods were typically used to measure a subject's performance. The magnitude of an electrical signal could also be summed (accumulated) easily enough. The longer the duration of a trial, the bigger the number became. This gave a single number or score for a trial. In the process, the variability over time was lost by using this single number to represent the string of data over time. An average does the same

thing: it divides a sum by a scalar value to give another, smaller scalar value. It smooths the variations in the data and suppresses information about the variability in the measures over time. While some information is thus lost, the average value is an economical and meaningful measure of overall performance.

If the signal could be compared to some reference signal, then a deviation signal could be generated. The deviation scores could be positive or negative, and since positive scores would cancel negative scores, it was common practice to always take the positive value of the deviation: its absolute value. This was easy to do electronically. Squaring a score achieves the same thing (converts all deviations to positive values), but the sum of squared values is a larger number than the corresponding sum of absolute values, and it is not as easy to do this directly, through electronics. If the data were digitized (which experimenters tried to avoid), the calculation of sums of squares could be done.

Since even the deviation scores fluctuated over the course of a trial, it was common to compute some sort of a value that indicated variability instead of constant bias or average error. The simplest of these was to simply note the maximum and minimum observed values (or deviations). The Root Mean Square (RMS) error was another popular measure. Variance and standard deviation computations were not easily accomplished in analogue systems. They required data reduction to get digitized values that could then be submitted to appropriate statistical analysis. If the mean or average value was zero, then the root mean square would be the standard deviation. For a non-zero mean, the RMS value is related to the square root of the sum of squared scores, one element of the standard deviation computation.

An average value reflects a measure of central tendency in a set of scores. If the scores are symmetrically distributed around the average value, then other measures of central tendency (such as the mode, and median) are the same value. When the distribution is not symmetric, then the mean is influenced more by extreme scores than the median will be. If a deviation score ($x=X-C$) is computed, it can be computed with reference to any particular value. If the deviation is computed with respect to the average value (i.e., we let C=mean), then it is known that the sum of deviations will be zero. For any other selected value of C, we would therefore expect the sum to be non-zero. A non-zero deviation score typically reflects a bias or constant error from the target value (C).

Variability can be similarly represented by a scalar value a number of ways. The range (maximum – minimum) is one such value. It is typically less reliable than other measures that use more of the information in a set of scores. The variance and standard deviation can be shown to be efficient statistics for variability. The average absolute value of errors is similar to the average squared error (variance), but it can be shown that the variance is a more efficient statistic, and on that grounds, it is preferred.

As an estimate, 1/6 of the range is approximately equal to the standard deviation (assuming 6 standard deviations encompasses approximately 99% of the variability in a set of data). This fact can often be used as a useful cross-check for computational reasonableness.

What these summary statistics (average and standard deviation) ignore is the time history of the error: the time trajectory. The time trajectory reflects the variation on a moment to moment or continuous basis. While a scalar value is economical, it filters the raw data and may hide the information of greatest importance to an instructor: what happens at a particular moment, rather than overall. Very different time trajectories can generate identical summary statistics. While the average and standard deviation are efficient statistics and more economical than the time trajectory, they may actually suppress important diagnostic information that would help in remediating skill deficiencies.

Methodology

Maneuvers typically begin by launching MS FS, putting it in the Pause mode (by pressing the P key), selecting the data recording function from the appropriate pull down menu, and then selecting the maneuver to be performed. For each maneuver, the trainee must begin a turn, climb, or descent from straight and level flight. The instructor takes MS FS out of pause mode by again pressing the P key. The instructor then waits for the student to attain and maintain (for about 5 seconds) straight and level flight.

The student pilot is told to begin as soon as the instructor / evaluator signals the maneuver should begin. The maneuver is initiated and the transition from straight and level to the desired maneuver criterion is accomplished (bank angle, turn rate, descent rate, climb rate, etc.) The maneuver ends when the aircraft is stabilized again in straight and level flight. By pressing the P key, MS FS is put in a pause mode: that stops the computer and allows the selection of the next maneuver to be performed.

During the maneuver, selected performance parameters are sampled by the computer at one second intervals. Some time (less than three seconds) is allowed for the trainee to establish the maneuver before scoring data is extracted. This corresponds not only to the trainee's reaction time to the instructor's start command, but the lag in the aircraft's dynamic response to the pilot's input(s). The purpose of this is to eliminate the transition data. Scores will be determined after the maneuver is completed by comparing the trainee's performance of the maneuver to a criterion. For example, if the instruction was to maintain a given bank angle, then bank angle is compared to the criterion value. If instead, the student was asked to maintain a turn rate, then performance is compared to the turn rate criterion.

Scoring is accomplished by extracting the raw data (X) for a fixed set of variables (altitude, air speed, bank angle, pitch angle, etc.) from the MS-FS program. These data are passed to a separate performance scoring program that will automatically establish values of deviations (x) from desired (X') values (maneuver criterion) and convert those deviations (x = (X-X')) to scores. A score is assigned based upon the magnitude of the deviation (how much the actual value varies from the desired value)

In general, a higher score is assigned for a larger deviation, since the scoring is based on the magnitude of the error between actual and desired value. The larger the score, the poorer the performance. The sampled scores could be summed over time to establish a raw score for the maneuver as a whole, but that is not recommended, for reasons discussed later. The merits and disadvantages of alternate scoring methods are treated in greater detail in a later section of this report.

The scores can be assigned in either of two ways. Small errors can be given a large value or they can be given a small value. At first, we assigned the best performance a score of 4 and worst performance a score of zero. While this works, it seems inconsistent with the scoring of error, where zero error is good. In golf, the low score wins. So, a low score is a good score. What this allows is an increasing score for an increasingly large error. While the scale is presently truncated at some upper level, additional values could be easily added. When the scale is inverted (low error = high score), one would have to use negative values to capture errors larger than the one assigned a zero value. This seemed odd. So we finally decided to go back to the concept that the low score is a winner and made zero error = zero points.

The sampled scores (X) and derived deviation scores (x) create two data streams or time trajectories for every measured parameter. The scoring process consists of tabulating the deviation score (x) time trajectory and comparing the magnitude of the deviation against criterial deviation levels. Reserving the number zero for the case where there is no error at all, scores are assigned according to which of four such criterion levels (if any) have been crossed. If the pilot's performance leads to deviations below (smaller than) the tightest criterion value, then the best score (1) is assigned. If at any time during the maneuver, the pilot exceeds this value, then a higher score will be assigned. The score of 2 will be assigned, unless sometime during the maneuver, the deviation is larger than the second level criterial value. In that case, a 3 would be assigned, unless sometime during the maneuver, the deviation was larger than the third specified criterial value. A 4 would be assigned, unless sometime during the maneuver, the pilot's performance led to a deviation greater than the fourth criterial level. In that case, the pilot's score would be 5 for the maneuver: the PTS value would have been exceeded.

This stratification of the error into levels can be varied, but there should be some rationale for establishing these criterial levels. For development, they were set somewhat arbitrarily. In practice, more research is needed to determine how best to set the four cutoff levels. The parameter values used for scoring are documented in Appendix A. In the case of altitude scores, a reasonable rationale can be offered, as explained later.

Most maneuvers require holding more than one parameter within specified limits (e.g. turning without losing altitude). Therefore, each of the required parameters will be scored in this same fashion, since each will have its own prescribed criterial values. So each maneuver will generate a profile of scores, not just a single value. However, the maximum score for the primary measure for each maneuver in each sequence determines the assigned score for the maneuver. Other segments of the scored data (such as areas where scores of 5 were obtained) will be

available to determine areas where more training is needed. A frequency count of the scores is also available: how many (what % of the time) a particular score was obtained.

To get a single value for an overall score, some combination of the profile scores is required. Obviously, the best possible score is to get a "0" (or practically speaking, a "1") in all the parameters. The problem arises when one or more of the scores is not a "1." As a first step in devising a composite score, the parameters associated with a maneuver should be ranked in terms of their contribution to optimum performance. This can vary from one maneuver to the next. For example, early in training, learning to hold bank angle constant is more important than keeping the turn rate constant. Later in training, it is important to keep the rate of turn constant, even if bank angle has to be adjusted to do that.

Once the parameters have been ranked, one could weight them. That is the part we have not yet done. The question is what to use as the basis for weighting one parameter more or less important than some other parameter. If there was some external criterion for what constitutes the best maneuver, then multiple regression techniques could be used to derive weights that best predict the criterion variable. However, no such external criterion measure exists.

An alternate approach is suggested, but its implementation has not been attempted. The question is whether the parameters are of equal or unequal importance. If all five parameters were of equal importance, then each parameter would have the same weight. Say the total for the weights is 100, and these 100 points are to be allocated to the parameters. If all are equally important, then 20 points should be assigned to each of the five. However, if one is more important than the others, then it is assigned more points, which means points have to be taken away from the other parameters. The number of points assigned then reflects the importance of the parameter with respect to the other parameters.

This weighting scheme is arbitrary and only reflects the subjective opinion of those providing the weights. Until some measurable external criterion is defined, this is the only feasible approach for constructing a composite score from the scoring profile. Without the subjective allocation of weights, the scoring algorithm cannot be constructed. As part of our validation study, we attempted to get preliminary values for these weights. The training objective will to some degree influence the weight a parameter receives in deriving the index (single value) for scoring how well the maneuver was performed. How to set these weightings appropriately will need to be the subject of subsequent research.

Much of what is done in mathematics requires a single valued function: a criterion (dependent) variable that is expressed as a function of one or more factors or independent variables. If we do not have such a function, then we may need to construct one. Measures of merit or objective functions in operations research are examples of doing this, so tradeoff analyses and optimization can be accomplished. A set of variables are combined into a composite

index by weighting and then summing the contributions of the individual variables. Analysis of variance is also an example of a linear, additive model of this sort.

Many research tasks have assumed that a single criterion variable is sufficient to measure desired performance or serve to indicate superior performance in some particular task. However, there are a number of tasks where multiple variables have to be maintained, some to greater or lesser precision than others. This of course requires time-sharing on the part of the performer, to assure that appropriate control inputs are made to achieve the required criterion levels on all critical or important variables concurrently.

Also, the dynamics of the system influence the nature of the task. In many laboratory tracking tasks, these dynamics have been simplified and do not represent the complexities of the actual task as it is performed in the system operating context. For example, in aircraft, changes in speed will affect altitude. Changes in pitch affect speed. In a turn, airspeed is lost as well as altitude unless other control inputs are supplied to compensate.
The dynamics of the flight vehicle are cross-coupled: actions designed to control one variable affect other variables as well. To achieve the desired outcome, mulitple variables may have to be manipulated concurrently and in appropriate proportion to one another. Those relationships may not be constant either. They may vary over time or under differing environmental conditions.

For example, air density affects aircraft performance. On a cold, crisp, winter day, the aircraft is much more responsive than it will be on a hot, humid, summer day. An aircraft taking off at a higher elevation will have a longer takeoff roll than one operating at a lower elevation, simply because air density changes as altitude (and elevation) increase. These factors have to be learned and anticipated by the pilot.

Every geographic area seems to have its own unique weather patterns. In Florida, large thunderstorms occur daily, in some seasons, in the mid- to late afternoon. Storms are regular enough to be anticipated: you cannot claim surprise if you live in the area long. By contrast, in Arizona, the weather is "severe clear" most of the time. However, when storms come, the are typically quite severe. The Midwest has changing weather patterns than can surprise the inattentive, and offer some unique problems not seen in the other two areas: ice. Ice buildups not only create a heavier (less flyable) aircraft, the ice buildup can also change the dynamics of the airfoil that creates lift, counteracting the weight.

Mountains, big or small, change the airflow in their vicinity. Pilots have to anticipate up and downdrafts in the area of mountains and adjust their flight path accordingly. Also, for tall mountains, they need to plan ahead and climb to a safe altitude before reaching the barriers, so they can safely pass over, through, or around these natural barriers to flight. So there are both short term and long term dynamic changes that affect the flying task, and students who are expected to fly in more than one climate, season, or locale will have to learn how to accommodate these changes as they become skilled and proficient pilots.

In order to create a proficient skill, the student must first learn the skill (how to execute a maneuver) and then develop the proficiency to do it well. In the learning of the skill, it may be necessary to focus on different aspects of the task and change that focus as skill progresses. For example, in teaching turns, coordination of activities may be ignored at first, simply to assure that the basic control inputs are learned and that the student learns to anticipate the effects.

At this stage, the instructor may ask the student to achieve a fixed bank angle, allowing other variables to change as they may. Later, the instructor will allow bank angle to vary in order to achieve a constant turn rate. Sometime between these extremes, the student must also learn not to let the nose drop during a turn (by changing pitch, by adding power, or by some combination of these actions). Also, the student will need to learn how to add an appropriate amount of rudder to correct skids and slips during a turn.

Consequently, during the course of training, the criterion variable(s) and their relative importance may change. The scoring system proposed should therefore be flexible enough to reflect these changes so what gets emphasized in scoring matches the training objective at this stage of skill and proficiency development. It is not clear how to set the values, it is only clear that they probably need to be changed as training progresses. Empirical study will be needed to determine how to set the scoring weights.

While the FAA's PTS standards were used as the basis for setting the upper limits on altitude requirements in our scoring system, it might be useful to provide a rationale for the 200 foot value used. VFR traffic is separated from IFR traffic by 1000 feet in the following fashion. First, all IFR traffic is assigned to a particular altitude, which may be an even or odd number: 2,000 feet, 4,000 feet, 6,000 feet, etc. or 3,000 feet, 5000 feet, 7,000 feet. By contrast, VFR traffic is assigned to the 500 foot level between these altitudes: 2500, 4500, 6500, etc., or 3500, 5500, 7500, etc. The even numbered altitudes correspond to traffic headed roughly West (180-359 degrees) and the odd numbers to East bound traffic (0-179 degrees).

As a consequence, aircraft operating at exactly the right altitude should clear each other by 500 feet, no matter what kind of traffic they are or in which direction they are traveling. Since aircraft will not be exactly at the assigned altitude, due to instrument errors, weather (pressure) variations, and pilot error, an aircraft might be higher or lower than that value by some amount. Clearly, there should be a high probability that the two aircraft will not collide. Setting 200 (instead of 250) feet as the 99% confidence limit for deviations about the assigned altitude means that two aircraft at the limit of the allowable deviation would pass within 50 feet of each other. That is close enough for anyone!

If we take 200 feet as the three sigma value for allowable altitude errors, then the one sigma value is 67.7 feet and the two sigma value is 133.33 feet. If the standard deviation of a pilot's altitude control errors is 67.77 feet or less,

then we have reason to believe that individual can meet the PTS standards 99% of the time. Also, the pilot will keep the airplane within 150 feet more than 95% of the time, or within 100 feet around 60% of the time. The statistics give us some notion of how often or how likely an adverse altitude excursion would be.

Results

Since there is no absolute external criterion, validation of the scoring method really means comparing the instructor's scores with the automated scores. The instructor's ratings of pilot performance are what is used currently to grade or evaluate student performance of maneuvers. Typically the score is based on a percentage value and is assigned for an entire lesson or for some particular learning objective.

In our case, something more definitive was desired, something closer to what the automated scoring system did. Consequently, we wanted instructors to use the same 1-5 score being assigned by the scoring system, and simply assign a number to each of the parameters, reflecting how well they thought the student had done in performing the maneuver.

While we had planned to collect data on a number of students performing with a number of instructors, the debugging of the software took longer than anticipated, and we lost our window of opportunity. The academic year ended without being able to run the study and collect the data.

Our plan was to have every instructor grade every variable / measure on every maneuver that each of their students performed. Students would be nested within instructors in this design; it is not fully crossed. That is, instructors would not evaluate all students, only their own students (three per instructor). This allows instructor scores to be compared with numerically derived scores on all variables. The following two null hypotheses could tested:

1. There is no difference between instructor ratings and the numerically derived scores on the primary measure for a maneuver.
2. There is no difference between instructor ratings and the numerically derived scores on any of the associated parameters for a maneuver.

The second hypothesis is actually a set of hypotheses, since it applies to each of the secondary measures for a maneuver.

The alternative hypotheses are:

1. The instructor ratings differ significantly from the derived scores on the primary variable for a maneuver, indicating that the two scores may not be measuring the same thing.

2. The instructor ratings differ significantly from the derived scores, indicating the instructors cannot watch all parameters equally well, if for no other reason, they have trouble dividing their attention.

This rest of this section discusses the measures that can be collected from MS FS. It also explains the weights chosen in this preliminary study and rationalizes the choices.

MS FS runs under Windows. While it can run under either Windows 95 or Windows 98, we did not have much success running it under Windows NT, although it should do so. While MS FS can be run from keyboard commands, that is not very realistic, and it is recommended that MS FS be used with some kind of yoke and rudder pedal kit. Several vendors products are available, and problems are typically encountered getting these to work with your particular computer, so be advised and seek appropriate help.

Once everything seems to be working reasonably (the yoke changes, throttle changes, and rudder pedal inputs seem to have an influence on the MS FS operation), then some degree of calibration is warranted. There are at least two considerations. First, the MS FS package has built-in routines for yoke and throttle calibrations. Second, there are adjustable sensitivities and run-time set-up features that will affect how well the simulated aircraft responds to control inputs.

The built-in MS FS calibration routines simply scale the yoke and throttle inputs so that the software and hardware work well together. The full yoke deflection points are identified to the MS FS software package so it can interpret and calculate changes in yoke position from one end of the scale to the other. The same is true for the throttle adjustments. Calibration does not make these "accurate" in any sense. It just assures that the MS FS software knows how to interpret the signals it receives from the yoke and throttle.

The sensitivity of aircraft controls can be adjusted, but not truly calibrated. These changes do affect the handling qualities of the simulated aircraft, but there is no simple, scientific way to assure that they represent any particular aircraft or operating condition. It is important to set these so they are not too sensitive, because that not only makes the aircraft hard to control, it makes it unlikely that you can satisfy the requirements of the data collection system: starting and ending with a stable flight attitude.

If a trial takes too long, or the data are contaminated, the scoring software may not operate properly. When run time errors are encountered, MS-FS may still operate, but the scoring software terminates. It is then necessary to start over: shut down MS-FS and restart the scoring software per the directions in Appendix B. This can make for a very long day! So be sure the MS-FS software is set up and operating in a manner that gives the test subject a reasonable flying task rather than a super challenging one, at least while they are learning. Otherwise, the scoring system will only work with your very best pilots, if at all.

Also, in MS-FS there is an inherent tradeoff between updating the visual scenery and updating the equations of motion that represent aircraft dynamics. The higher resolution scene detail requires more computer speed and memory than lower resolution scene detail. Often, to get realistic updates of scenery, one has to sacrifice dynamic aircraft response. To get more dynamic aircraft response, one has to live with less detailed scenery. Since our emphasis is in performing maneuvers, we recommend not using run set up features that use high-resolution scenery.

When MS FS executes, it does so on a cyclic basis. Once every cycle, it will go out and check whether some user-supplied routine needs to be executed. If the user has put an executable load module in the right subdirectory, then MS FS will execute that module whenever it comes to this point in its operating cycle. That allows third party software packages to interact with MS FS at run-time.

The product of this study effort is a set of 3 ½ inch diskettes that will self-install the scoring system on a PC. It will create the necessary links to MS-FS for run-time execution, collection of data, and storing the scores for post-run retrieval. From experience, we have learned that this self-install routine will not be successful in every case, due to idiosyncrasies in a particular PC, its hardware, its software, and the configuration of both.

As an alternative, we have had to install the development software (Visual Basic) on some systems in order to get the scoring package to work successfully. While this takes more expertise and space on the machine, it has proven successful in those cases where the self-install did not work correctly.

The basic run-time data collection routine is a third party software module that is free ware provided on the internet. This routine is called once each operating cycle and examines a set of memory locations, reads their content, and writes those values to a data file on the hard disk. These data include six variables of interest to us: 1) airspeed, 2) altitude, 3) heading, 4) bank angle, 5) vertical velocity , and 6) turn rate.

In earlier versions of this software, the data were captured as binary digits and then had to be converted to ASCII characters afterward. The present version supplies converted data as the raw data. That means the numbers are interpretable values when saved in the file generated by this third party software routine. While considered "freeware" most shareware of this sort typically requests that users provide some nominal payment for continued use of the software. The scoring system described in this report is dependent on use of this third party software and will not operate without it.

It is also possible to include routines in this same subdirectory that influence the operation of MS FS during run time. So, at least in theory, one could examine a run result, interpret its meaning, and then alter some display in MS FS that the student would see. MS FS would not have to be modified to do this, but the speed of execution would be slowed by the amount of time taken to execute these added routines. Clearly, a more responsive approach would be to embed such code into MS FS itself. However, that requires the cooperation of Microsoft program developers.

The scoring system was built as a Microsoft Windows application. It is designed to be installed separately, but it operates in conjunction with an already installed MS FS program and can be operated from within the MS FS program (once it is launched and operating), by selecting the appropriate pull down menu (MODULES). Select the Flight Data Recorder option. This action will then launch the installed scoring software and bring up a dialogue box for operating the scoring system. More complete instructions are provided in Appendix A.

The data are actually stored in a Microsoft Access data file. The scoring system interacts with this file to present the final scores for a particular maneuver. The user can enter data about the student, which is then used to create a unique record for each data collection run – identifying who did which maneuver under what circumstances.

The scoring system gives the final score for a subject, but the entire data stream has been captured in the Access data file, should anyone want to go back and look at that data. No additional analyses have been performed on those records during this study. Provisions were simply made to provide users with additional data should they wish to do their own analysis after a study, in the conventional off-line fashion – using whatever statistical data analysis package they may wish to employ.

Conclusion

A window-based pilot scoring system was developed and implemented with Microsoft Flight Simulator (MS-FS), a low-cost, commercial off the shelf (COTS) software package. While originally designed for the game market, improvements made to this product have made it suitable as a training aid for various aspects of flight instruction.

Subsequent studies should examine how instructors and students use the information provided by the numerical scoring system. Just having some surface agreement between instructors and the scoring system does not mean it is being used to advantage for achieving the instructional objective. It only means it is an acceptable surrogate for evaluations.

While Microsoft Flight Simulator (MS-FS) is an acceptable flight simulation for initial use at home, it is not certified or approved by the FAA for flight instruction. On the other hand, FlitePro by Jeppesen is an approved Aircrew Training Device (ATD) for the Personal Computer (PC). This PCATD is especially useful for staying proficient at instrument flying skills. It is recommended that the next stage of development examine the use of this product and how a scoring system might be embedded in it. In order to get useful data from flight, it is also recommended that a copy of FlitMap and an ETAK GPS be acquired, in order to track actual aircraft flight. These products may afford the possibility of getting a low-cost airborne instrumentation system that could then be used with the scoring system developed here.

Appendix A

Parameters for Scoring Pilot Maneuvers

Instructor Name ____________________ **Trainee** ____________________

Pilot rating_________(i.e. none, student, private, etc.) Date______________

Actual flight hours: C-150/152_____ **C-172/182**____ **Complex** ________

Other flight time (specify)____

Flight Simulator time (estimated)_______**Sim type** ________________

INSTRUCTOR GRADING CRITERIA FOR PARAMETERS FOR FLIGHT MANEUVERS BY JOSEPH L. VOGEL (Modified 06/03/99)

The following parameters address the flight training maneuvers that are to be tested and graded by the flight instructor. Parameters are taken from FAA-S-8081-14, PRIVATE PILOT PRACTICAL TEST STANDARDS FOR AIRPLANE dated May 1995. Additionally, parameters are taken from several of the flight training publications such as AC 61-21, FLIGHT TRAINING HANDBOOK, and AC 61-23, PILOT'S HANDBOOK OF AERONAUTICAL KNOWLEDGE and the AIM (Airman's Information Manual.) Reference was also made to commercially available publications and from the personal flight instructing experience of the writer.

The purpose of this exercise is to validate a program of computer scoring of a pilot trainee's performance while "flying" a series of maneuvers on a low cost flight simulator. Microsoft's Flight Simulator (MSFS) is being used for this experiment. Validation will be accomplished by having the instructor/evaluator visually observe the maneuver and mark the observed score in the places provided in the accompanying form. The instructor/evaluator will have control of the computer for starting and stopping the automatic scoring of each maneuver and for evaluation by the computer after the maneuvers are flown. Control is accomplished by utilizing pull-down menus and the mouse controller.

The maneuver is initiated and completed when the instructor announces it to the trainee. For each maneuver, the trainee must begin to establish the parameters as soon as the instructor/evaluator signals the beginning of the maneuver. The computer is programmed to detect the establishment of the maneuver and record adherence to parameters after a steady state is established.

The instructor will visually observe and score each maneuver as it is being "flown" and will record those scores on the form in the places provided. During each maneuver, the instructor will wait until the maneuver is established and grade the maneuver by observing the maximum deviation from the ideal and score according to the criteria provided. For instance, grading airspeed deviations during a climb would entail waiting until the trainee stabilizes the airspeed, observing the deviation from the established value and placing a check mark next to the value observed (i.e.) if the trainee exceeds the established value (75 knots) by more than 5 knots but less than 6 knots the instructor would place a check mark beside the "Score: 1 _____. If the trainee, later in the maneuver, exceeds a greater value such as more than 8 knots but less than 10 knots, a check mark will be placed next to "Score 3 ____." The lower score will be the one compared to the computer score for the purpose of this experiment.

THE CLIMB (CONSTANT AIRSPEED)
Constant airspeed climb: Target airspeed 75 knots
(Parameters based on a Cessna 182 aircraft.)

1. Trainee will establish V_y (Best rate of climb = 75 knots) airspeed within plus or minus 10 knots.
   a. Plus 5, minus 5 knots — Score 1 ____
   b. Plus 6, minus 6 knots — Score 2 ____
   c. Plus 8, minus 8 knots — Score 3 ____
   d. Plus 10, minus 10 knots — Score 4 ____
   e. Outside of parameters — Score 5 ____

2. Trainee will maintain given heading plus or minus 20 degrees.
   a. Plus 5, minus 5 degrees — Score 1 ____
   b. Plus 10, minus 10 degrees — Score 2 ____
   c. Plus 15, minus 15 degrees — Score 3 ____
   d. Plus 20 minus 20 degrees — Score 4 ____
   e. Outside parameters — Score 5 ____

3. Once airspeed is established, trainee will maintain established pitch attitude plus or minus 8 degrees.
   a. Plus 3, minus 3 degrees — Score 1 ____
   b. Plus 5, minus 5 degrees — Score 2 ____
   c. Plus 7, minus 7 degrees — Score 3 ____
   d. Plus 8, minus 8 degrees — Score 4 ____
   e. Outside parameters — Score 5 ____

4. Once airspeed is established, trainee will not vary vertical speed by more than plus or minus 500 feet per minute (FPM).
   a. Plus 80, minus 80 FPM — Score 1 ____
   b. Plus 200, minus 200 FPM — Score 2 ____
   c. Plus 300, minus 300 FPM — Score 3 ____
   d. Plus 500, minus 500 FPM — Score 4 ____
   e. Over 500 FPM — Score 5 ____

5. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.
   a. Plus 50 minus 50 feet — Score 1 ____
   b. Plus 100 minus 100 feet — Score 2 ____
   c. Plus 150 minus 150 feet — Score 3 ____
   d. Plus 200 minus 200 feet — Score 4 ____
   e. More than 200 feet — Score 5 ____

THE DESCENT (POWER OFF)

Constant airspeed descent: Target airspeed 75 knots

(Parameters based on a Cessna 182 aircraft)

Trainee will reduce power to idle, hold the aircraft level and establish the best glide airspeed then establish the proper glide angle to maintain that airspeed. Target airspeed is 75 knots.

1. Once airspeed (75 knots) is established, trainee will maintain that airspeed, plus or minus 10 knots.

a. Plus 5, minus 5 knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8 minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside parameters	Score 5 ____

2. Once airspeed is established, trainee will maintain established pitch attitude plus or minus 10 degrees.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 6 minus 6 degrees	Score 2 ____
c. Plus 8, minus 8 degrees	Score 3 ____
d. Plus 10, minus 10 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

3. Trainee will maintain given heading plus or minus 20 degrees.

a. Plus 8, minus 8 degrees	Score 1 ____
b. Plus 10, minus 10 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20, minus 20 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

4. Once airspeed is established, trainee will not vary vertical speed by more than plus or minus 500 feet per minute (FPM).

a. Plus 200, minus 200 FPM	Score 1 ____
b. Plus 300, minus 300 FPM	Score 2 ____
c. Plus 400, minus 400 FPM	Score 3 ____
d. Plus 500, minus 500 FPM	Score 4 ____
e. Over 500 FPM	Score 5 ____

5. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

TURNS TO HEADINGS
Target airspeed 100 knots
(Parameters based on a Cessna 182 aircraft)

1. Trainee will establish an angle of bank for a medium banked turn. A medium banked turn is defined as one in which the bank angle is maintained to achieve a rate of turn at 3 degrees per second.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 6, minus 6 degrees	Score 2 ____
c. Plus 7, minus 7 degrees	Score 3 ____
d. Plus 8, minus 8 degrees	Score 4 ____
e. Over plus or minus 8 degrees bank	Score 5 ____

2. Once bank is established, trainee will maintain altitude within plus or minus 200 feet.

a. Plus 50, minus 50 feet	Score 1 ____
b. Plus 100, minus 100 feet	Score 2 ____
c. Plus 150, minus 150 feet	Score 3 ____
d. Plus 200, minus 200 feet	Score 4 ____
e. Outside parameters	Score 5 ____

3. Trainee will maintain cruise airspeed within plus or minus 10 knots.

a. Plus 3, minus 3 Knots	Score 1 ____
b. Plus 7, minus 7 knots	Score 2 ____
c. Plus 9, minus 9 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

4. Trainee will roll out of turn on assigned heading plus or minus 20 degrees. Trainee will hold that heading.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 8 minus 8 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20, minus 20 degrees	Score 4 ____
e. Outside of parameters	Score 5 ____

STRAIGHT AND LEVEL FLIGHT
Target cruise airspeed, 100 knots
(Parameters based on a Cessna 182 aircraft)

1. Straight and level flight will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

2. Trainee will maintain given heading plus or minus 20 degrees.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 8, minus 8 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20 minus 20 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

3. Trainee will establish cruise airspeed of 100 knots. Cruise airspeed will be maintained within plus or minus 10 knots.

a. Plus 5, minus 5 Knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8, minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

4. Once airspeed, altitude and heading is established, trainee will not vary vertical speed by more than plus or minus 500 feet per minute (FPM).

a. Plus 100, minus 100 FPM	Score 1 ____
b. Plus 200, minus 200 FPM	Score 2 ____
c. Plus 300, minus 300 FPM	Score 3 ____
d. Plus 400, minus 400 FPM	Score 4 ____
e. Over 500 FPM	Score 5 ____

CONSTANT AIRSPEED CLIMBING TURN

Target Airspeed 75 knots

(Parameters based on a Cessna 182 aircraft.)

NOTE: For this maneuver, trainee will simultaneously establish climb and bank attitude to maintain a constant airspeed, constant turn rate climbing turn. Assigned level off altitude will normally be 500 feet above the altitude the maneuver was started.

1. Trainee will establish V_y (Best rate of climb = 75 knots) and maintain airspeed within plus or minus 10 knots.

a. Plus 5, minus 5 knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8, minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

2. Trainee will establish an angle of bank for a medium banked turn. A medium banked turn is defined as one in which the bank angle is maintained to achieve a rate of turn at 3 degrees per second.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 7, minus 7 degrees	Score 2 ____
c. Plus 9, minus 9 degrees	Score 3 ____
d. Plus 8, minus 8 degrees	Score 4 ____
e. Over plus or minus 8 degrees bank	Score 5 ____

3. Once airspeed is established, trainee will not vary vertical speed by more than plus or minus 500 feet per minute (FPM).

a. Plus 80, minus 80 FPM	Score 1 ____
b. Plus 200, minus 200 FPM	Score 2 ____
c. Plus 300, minus 300 FPM	Score 3 ____
d. Plus 500, minus 500 FPM	Score 4 ____
e. Over 500 FPM	Score 5 ____

4. Trainee will roll out of turn on assigned heading plus or minus 20 degrees. Trainee will hold that heading.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 10 minus 10 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20, minus 20 degrees	Score 4 ____
e. Outside of parameters	Score 5 ____

NOTE: Assigned level off altitude will normally be 500 feet above the altitude the maneuver was started.

5. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

CONSTANT AIRSPEED DESCENDING TURN

Constant airspeed descent: Target airspeed 75 knots

(Parameters based on a Cessna 182 aircraft)

Trainee will reduce power to idle, hold the aircraft level and establish the best glide airspeed then simultaneously establish the proper glide and bank angle to maintain that airspeed and rate of turn. Target airspeed is 75 knots. Rate of turn target is 3 degrees per second.

1. Once airspeed is established, trainee will maintain that airspeed, plus or minus 10 knots.
   a. Plus 5, minus 5 knots Score 1 ____
   b. Plus 6, minus 6 knots Score 2 ____
   c. Plus 8 minus 8 knots Score 3 ____
   d. Plus 10, minus 10 knots Score 4 ____
   e. Outside parameters Score 5 ____

2. Trainee will establish an angle of bank for a medium banked turn. A medium banked turn is defined as one in which the bank angle is maintained to achieve a rate of turn at 3 degrees per second.
   a. Plus 5, minus 5 degrees Score 1 ____
   b. Plus 7, minus 7 degrees Score 2 ____
   c. Plus 9, minus 9 degrees Score 3 ____
   d. Plus 8, minus 8 degrees Score 4 ____
   e. Over plus or minus 8 degrees bank Score 5 ____

3. Once airspeed and bank angle is established, trainee will not vary vertical speed by more than plus or minus 500 feet per minute (FPM).
   a. Plus 80, minus 80 FPM Score 1 ____
   b. Plus 200, minus 200 FPM Score 2 ____
   c. Plus 300, minus 300 FPM Score 3 ____
   d. Plus 400, minus 400 FPM Score 4 ____
   e. Over 500 FPM Score 5 ____

4. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.
   a. Plus 50 minus 50 feet Score 1 ____
   b. Plus 100 minus 100 feet Score 2 ____
   c. Plus 150 minus 150 feet Score 3 ____
   d. Plus 200 minus 200 feet Score 4 ____
   e. More than 200 feet Score 5 ____

5. Trainee will roll out of turn on assigned heading plus or minus 20 degrees. Trainee will hold that heading.
   a. Plus 5, minus 5 degrees Score 1 ____
   b. Plus 10 minus 10 degrees Score 2 ____
   c. Plus 15, minus 15 degrees Score 3 ____
   d. Plus 20, minus 20 degrees Score 4 ____
   e. Outside of parameters Score 5 ____

THE CLIMB (CONSTANT RATE OF CLIMB)
Target Airspeed 75 knots, target rate of climb, 500 feet per minute
(Parameters based on a Cessna 182 aircraft.)

1. Trainee will establish V_y (Best rate of climb = 75 knots) airspeed within plus or minus 10 knots.

a. Plus 5, minus 5 knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8, minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

2. Trainee will maintain given heading plus or minus 20 degrees.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 10, minus 10 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20 minus 20 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

3. Once airspeed is established, trainee will maintain established pitch attitude to hold 500 feet per minute rate of climb plus or minus 8 degrees.

a. Plus 3, minus 3 degrees	Score 1 ____
b. Plus 5, minus 5 degrees	Score 2 ____
c. Plus 7, minus 7 degrees	Score 3 ____
d. Plus 8, minus 8 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

4. Once airspeed is established, trainee will maintain vertical speed and not vary by more than plus or minus 200 feet per minute (FPM).

a. Plus 50, minus 50 FPM	Score 1 ____
b. Plus 100, minus 100 FPM	Score 2 ____
c. Plus 150, minus 150 FPM	Score 3 ____
d. Plus 200, minus 200 FPM	Score 4 ____
e. Over 200 FPM	Score 5 ____

5. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

CONSTANT RATE OF CLIMB - CLIMBING TURN

Target Airspeed 75 knots, target rate of climb, 500 feet per minute.

(Parameters based on a Cessna 182 aircraft.)

NOTE: For this maneuver, trainee will simultaneously increase power and establish climb and bank attitude to maintain a constant airspeed, constant turn rate and constant rate of climb climbing turn. Power setting will vary as needed.

1. Trainee will establish V_y (Best rate of climb = 75 knots) airspeed within plus or minus 10 knots.

a. Plus 5, minus 5 knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8, minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

2. Trainee will establish an angle of bank for a medium banked turn. A medium banked turn is defined as one in which the bank angle is maintained to achieve a rate of turn at 3 degrees per second.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 7, minus 7 degrees	Score 2 ____
c. Plus 9, minus 9 degrees	Score 3 ____
d. Plus 8, minus 8 degrees	Score 4 ____
e. Over plus or minus 8 degrees bank	Score 5 ____

3. Once airspeed is established, trainee will maintain established pitch attitude to hold 500 feet per minute rate of climb plus or minus 8 degrees.

a. Plus 3, minus 3 degrees	Score 1 ____
b. Plus 5, minus 5 degrees	Score 2 ____
c. Plus 7, minus 7 degrees	Score 3 ____
d. Plus 8, minus 8 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

4. Once airspeed is established, trainee will maintain vertical speed and not vary by more than plus or minus 200 feet per minute (FPM).

a. Plus 50, minus 50 FPM	Score 1 ____
b. Plus 100, minus 100 FPM	Score 2 ____
c. Plus 150, minus 150 FPM	Score 3 ____
d. Plus 200, minus 200 FPM	Score 4 ____
e. Over 200 FPM	Score 5 ____

5. Trainee will roll out of turn on assigned heading plus or minus 20 degrees.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 10 minus 10 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20, minus 20 degrees	Score 4 ____
e. Outside of parameters	Score 5 ____

5. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

CONSTANT AIRSPEED, CONSTANT RATE OF DESCENT
Target airspeed 75 knots, target rate of descent, 500 feet per minute
(Parameters based on a Cessna 182 aircraft)

Trainee will reduce power, hold the aircraft level and establish the target glide airspeed and target rate of descent then establish the proper glide angle and power setting to maintain those targets.

1. Once airspeed is established, trainee will maintain that airspeed, plus or minus 10 knots.

a. Plus 5, minus 5 knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8 minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside parameters	Score 5 ____

2. Once airspeed and rate of descent is established, trainee will maintain established pitch attitude plus or minus 10 degrees.

a. Plus 5, minus 5 degrees	Score 1 ____
b. Plus 6 minus 6 degrees	Score 2 ____
c. Plus 8, minus 8 degrees	Score 3 ____
d. Plus 10, minus 10 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

3. Once airspeed is established, trainee will maintain vertical speed and not vary by more than plus or minus 200 feet per minute.

a. Plus 50, minus 50 FPM	Score 1 ____
b. Plus 100, minus 100 FPM	Score 2 ____
c. Plus 150, minus 150 FPM	Score 3 ____
d. Plus 200, minus 200 FPM	Score 4 ____
e. Over 200 FPM	Score 5 ____

4. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

CONSTANT RATE OF DESCENT - CONSTANT AIRSPEED DESCENDING TURN

Target airspeed 75 knots, target rate of descent: 500 feet per minute

(Parameters based on a Cessna 182 aircraft)

Trainee will reduce power, hold the aircraft level and establish the target airspeed then establish the proper glide angle and power setting to maintain airspeed and rate of descent simultaneously. Target airspeed is 75 knots, target rate of descent is 500 feet per minute. Target rate of turn is 3 degrees per second.

1. Once airspeed is established, trainee will maintain that airspeed, plus or minus 10 knots.
   - a. Plus 5, minus 5 knots — Score 1 ____
   - b. Plus 6, minus 6 knots — Score 2 ____
   - c. Plus 8 minus 8 knots — Score 3 ____
   - d. Plus 10, minus 10 knots — Score 4 ____
   - e. Outside parameters — Score 5 ____

2. Once airspeed is established, trainee will maintain established pitch attitude plus or minus 10 degrees.
   - a. Plus 5, minus 5 degrees — Score 1 ____
   - b. Plus 6 minus 6 degrees — Score 2 ____
   - c. Plus 8, minus 8 degrees — Score 3 ____
   - d. Plus 10, minus 10 degrees — Score 4 ____
   - e. Outside parameters — Score 5 ____

3. Once airspeed is established, trainee will maintain vertical speed and not vary by more than plus or minus 200 feet per minute.
   - a. Plus 50, minus 50 FPM — Score 1 ____
   - b. Plus 100, minus 100 FPM — Score 2 ____
   - c. Plus 150, minus 150 FPM — Score 3 ____
   - d. Plus 200, minus 200 FPM — Score 4 ____
   - e. Over 200 FPM — Score 5 ____

4. Level off will be at assigned altitude plus or minus 200 feet. Once established at that altitude, trainee will maintain that altitude plus or minus 200 feet.
   - a. Plus 50 minus 50 feet — Score 1 ____
   - b. Plus 100 minus 100 feet — Score 2 ____
   - c. Plus 150 minus 150 feet — Score 3 ____
   - d. Plus 200 minus 200 feet — Score 4 ____
   - e. More than 200 feet — Score 5 ____

5. Trainee will roll out of turn on assigned heading plus or minus 20 degrees. Trainee will hold that heading.
   - a. Plus 5, minus 5 degrees — Score 1 ____
   - b. Plus 10 minus 10 degrees — Score 2 ____
   - c. Plus 15, minus 15 degrees — Score 3 ____
   - d. Plus 20, minus 20 degrees — Score 4 ____
   - e. Outside of parameters — Score 5 ____

CHANGE OF AIRSPEED IN STRAIGHT AND LEVEL FLIGHT
Target cruise airspeed, 100 knots, Target speed reduction, 30 knots
(Parameters based on a Cessna 182 aircraft.)

The objective of this maneuver is to maintain straight and level flight while reducing airspeed. This approximates transitioning from cruise to final approach speed. The maneuver begins with cruise flight and is initiated by a power reduction. The pitch attitude of the aircraft will have to be steadily increased to compensate for the loss of speed and to maintain altitude.

1. Once altitude is established, trainee will maintain altitude within plus or minus 200 feet.

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

2. Trainee will maintain given heading plus or minus 20 degrees.

a. Plus 5, minus 5 degrees	Score 0 ____
b. Plus 8, minus 8 degrees	Score 2 ____
c. Plus 15, minus 15 degrees	Score 3 ____
d. Plus 20 minus 20 degrees	Score 4 ____
e. Outside parameters	Score 5 ____

3. Trainee will establish cruise airspeed of 100 knots prior to beginning the maneuver. Cruise airspeed will be reduced to 70 knots by throttle reduction. Trainee will coordinate power and pitch to maintain altitude and airspeed. Grading criteria for maintaining 70 knots follows:

a. Plus 5, minus 5 Knots	Score 1 ____
b. Plus 6, minus 6 knots	Score 2 ____
c. Plus 8, minus 8 knots	Score 3 ____
d. Plus 10, minus 10 knots	Score 4 ____
e. Outside of parameters	Score 5 ____

4. Change of airspeed back to cruise speed, (100 knots) requires full throttle, and a gradual reduction of pitch attitude as the airspeed increases. The objective is to perform the recovery without loss or gain of altitude. Once recovery is begun, trainee will maintain altitude within plus or minus 200 feet. After cruise speed is obtained, throttle will be reduced to cruise setting. Scoring criteria for altitude deviation is:

a. Plus 50 minus 50 feet	Score 1 ____
b. Plus 100 minus 100 feet	Score 2 ____
c. Plus 150 minus 150 feet	Score 3 ____
d. Plus 200 minus 200 feet	Score 4 ____
e. More than 200 feet	Score 5 ____

Appendix B

OPERATING INSTRUCTIONS

AIRCRAFT TRAINING MANEUVERS SCORING SYSTEM

(Revised 06/11/99)

INTRODUCTION:

The Aircraft Training Maneuvers Scoring System is composed of Microsoft's Flight Simulator software, running on a Personal Computer and interfacing with a special program to provide scores for maneuvers normally done during a pilot training program. The maneuvers are described in an attachment to these instructions. [Note: Appendix A]

What follows is a step-by-step procedure for operating the software that runs the scoring system. This procedure presumes a modicum of computer operator experience since it will normally be used by a flight instructor to evaluate the performance of a student pilot. The student pilot will follow directions given by the flight instructor and will fly the maneuvers in the sequence determined by the program.

The instructor will visually observe the student pilot's performance of a maneuver and hand score it on the sheet provided. As a validation of the experiment, scores by the instructor pilot and the machine will be compared. The machine will sample the student pilot performance every one second during the assigned maneuver. At the end of the maneuver, the student or the instructor (Their choice) will pause the flight simulator (by depressing the letter "P" on the keyboard) and the instructor pilot will initiate computer scoring. When the scoring is completed and saved, the instructor will select the next maneuver and "un-pause" flight simulator.

In order to provide the scoring system with the proper setup before any maneuver, the student pilot must fly straight and level for at least 5 seconds prior to initiating the maneuver and must then stabilize in 5 seconds of straight and level after completing the maneuver. The scoring system can then sense the starting and stopping points of the maneuver and will score only those points in between. After all of the required maneuvers are "flown," scored and saved, results will be compiled and printed.

STEP-BY-STEP PROCEDURE

1. TURN ON THE COMPUTER AND ITS MONITOR ACCORDING TO THE OPERATING INSTRUCTIONS IN THE COMPUTER OPERATING MANUAL.
(Computer instructions may vary.)

2. If the Simulator Scoring System does not appear on the lower tool bar (e.g., the tool bar is not on the screen), then double click on the Simulator Scoring icon on your desk top. The Simulator Scoring window / startup screen should appear. Click the OK button. The startup window will now disappear leaving the Simulator Scoring Main Screen. DO NOT CLICK THE "QUIT" BUTTON UNTIL THE ABSOLUTE END OF A SESSION!! Set up the first maneuver (we suggest this be straight and level just to give the student / subject time to practice before running a series of maneuvers). To scroll through the list of available maneuvers that can be scored, click on the arrow to the right or left of the window that displays: "Maneuver" and look at the window above it until the manuever you want appears in that window. Now, minimize the window (click on the minus sign in the upper right corner of the window / dialogue box. "Simulator Scoring System" will now appear on the tool bar at the bottom of your display screen.

3. Click on "Flight Simulator" icon (Double click if required) or select it from the Program menu. As the software loads, a space scene appears (in some versions, you can skip this by clicking the OK button). Wait until the MS-FS Main Screen comes up (there are four action circles left mid-screen, a jet, and two button bars at the lower right: a) Fly Now?, and b) Exit).

4. Click on the "FLY NOW" button bar. (Instrument panel and outside scenery will appear.) THROTTLE, TRIM, AND WHEEL ON THE CONTROL YOKE WILL NOW OPERATE. You may also hear an engine sound if your computer is equipped with a sound card.

5. Press (.) (the period key on your computer keyboard) to release the parking brake. Unlike an airplane, the control wheel acts as a steering wheel on the "ground."

6. Add full throttle. Flaps are not required for takeoff.

7. After reaching 60 knots, rotate to approximately 15 degrees nose up. (Each bar on the attitude indicator is five degrees pitch).

8. Vary climb angle to maintain 75 knots climb speed. Climb to a particular altitude (e.g., some even thousands of feet – for easy reference) and stabilize straight and level flight. Press "P" on the keyboard to pause Flight Simulator now. (Later , you will again Press "P" on the keyboard to un-pause the program.)

9. To begin the scoring program, click on MODULES on the menu bar. The FLIGHT DATA RECORDER window will appear.

10. With the mouse, move the cursor (arrow) to: SETTINGS. Click on SETTINGS. The Flight Data window will appear.

11. Click on the ENABLE recorder box. A check mark should appear in the box .

12. Click on "OK."

13. Press "P" on the key board to continue (leaving the pause mode).

14. Fly the maneuver (At least 5 seconds must first be flown in a straight and level attitude for the program to recognize the start of a maneuver.) An additional 5 seconds of straight and level must be flown after completion of the maneuver so that the computer will recognize the end of that maneuver.

15. To stop recording the maneuver (So that it can be scored.) first press "P" to pause the Flight Simulator Program. The instrument panel will "freeze." Click on "MODULES" on the menu bar.

16. Place the cursor on FLIGHT RECORDER, and then click on SETTINGS. The FLIGHT DATA RECORDER WINDOW will appear.

17. In the FLIGHT DATA RECORDER window, click on the ENABLE recorder box. The check mark will disappear. This is an important step. The program will not score unless it is done. Click OK.

18. Pull the cursor arrow to the bottom of the screen. The task bar at the bottom of the screen will appear.

19. Click on SIMULATOR SCORING SYSTEM in the menu bar.

20. Click on "OK" in the opening screen (if it appears). SIMULATOR SCORING SYSTEM MAIN SCREEN will appear.

21. Select the pilot's name from those in the list by clicking on the (< >) arrows in the "PILOT" box.

22. Another pilot name can be added by clicking on "NEW PILOT." The "PILOTS" window will open. Follow the prompts to add a new pilot. Click on "SAVE" and then "CLOSE." [NOTE: This function may not always work properly in the delivered software; simply use one of the built in names.]

23. In the SIMULATOR SCORING SYSTEM MAIN SCREEN, insure that the desired pilot name is present and the maneuver flown appears in the "MANEUVER" box. (Scroll through the maneuvers by pressing the arrows for right or left scroll.)

24. Click on the "PROCESS FLIGHT" button. WAIT -- Listen for the 'beep" which indicates that the flight has been processed.

25. Press the SCORE command button. SCORING RESULTS screen will appear.

26. Click on the SCORE button. Chime sounds and scores are presented.

27. Click on the SAVE FLIGHT AND SUMMARY button.

DEVELOPMENT AND VALIDATION OF A PHYSIOLOGICALLY-BASED KINETIC MODEL OF PERFUSED LIVER FOR WATER SOLUBLE COMPOUNDS

Brent D. Foy
Assistant Professor
Department of Physics

Wright State University
Dayton, Ohio 45435

Final Report for:
Summer Research Extension Program
Wright Research Site

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC
and
Wright Research Site

March, 1999

DEVELOPMENT AND VALIDATION OF A PHYSIOLOGICALLY-BASED KINETIC MODEL OF PERFUSED LIVER FOR WATER SOLUBLE COMPOUNDS

Brent D. Foy
Assistant Professor
Department of Physics
Wright State University

Abstract

The appearance of bromosulfophthalein (BSP) and its glutathione conjugate in bile is a commonly used indicator of liver function. A more complete understanding of the kinetics of the physiological steps involved in BSP metabolism will enhance the utility of BSP as a test compound, and also provide a framework for the kinetic analysis of other toxins. Using an isolated perfused rat liver system with recirculating perfusion medium, the liver was exposed to perfusion medium containing 0, 0.25, 1, or 4% bovine serum albumin (BSA, w/v). In another series of experiments, the liver was perfused with a single pass of the perfusion medium, and four combinations of BSP and BSA. The modeling focus was on integrating protein binding kinetics and metabolism in a single model of perfused liver BSP kinetics. The results indicate that a strong binding interaction, beyond keeping the concentration of free chemical low due to a small equilibrium dissociation constant, can also reduce uptake by an organ due to the slow release of chemical from the protein during passage through the capillaries. With respect to metabolism, the presence of lower conjugation fractions at higher BSA concentrations was somewhat surprising since there is less likelihood that the metabolism process would be saturated when total BSP uptake and output in the bile was lower. This led to a hypothesis that at low intracellular BSP concentrations, biliary excretion of non-conjugated BSP is preferred over metabolism, but at higher intracellular BSP concentrations, biliary excretion of non-conjugated BSP is saturated leading to an increased rate of metabolism and excretion of conjugated BSP. The implication of these kinetic findings when extrapolating to other doses and *in vivo* situations is discussed.

DEVELOPMENT AND VALIDATION OF A PHYSIOLOGICALLY-BASED KINETIC MODEL OF PERFUSED LIVER FOR WATER SOLUBLE COMPOUNDS

Brent D. Foy

Introduction

Identifying and quantifying the cellular and physiological processes disrupted by a toxic compound is a critical step in the process of determining acceptable exposure limits and medical responses in the event of acute or chronic exposure. A standard way to probe cell-level physiological changes caused by a toxin is to quantify changes in some normal process for an organ, such as secretion of a protein or oxygen consumption. The uptake and metabolism of an easily monitored test compound, namely bromosulfophthalein (BSP), can also serve to probe toxic effects. BSP kinetics, such as its rate of disappearance from blood, has been used routinely to evaluate liver function. In such studies, typically the total BSP excretion over some period of time, or the total uptake of BSP into the liver is monitored. In recent studies, the analysis has been extended to evaluate the unidirectional fluxes across membranes using more complete models of BSP kinetics (Gartner et. al., 1997).

Through recent modeling and experimental investigations of the isolated perfused rat liver (IPRL), we have developed a comprehensive model of BSP metabolism and kinetics (Frazier *et al*, 1998; Foy *et al*, 1999). Since BSP undergoes several processes in its interaction with the liver—including transport through the sinusoidal membrane, conjugation with glutathione (forming BSP-GSH), and secretion of both BSP and BPS-GSH into the bile—it serves as a useful compound to analyze the disruption of physiological/biochemical processes by a toxin. With this complex system, a single experimental finding, such as reduced appearance of conjugated BSP in the bile, can have a number of causes. The goal is to refine the BSP analysis to the point where we can use it to rapidly identify the physiological/biochemical changes in the liver produced by a toxin.

A particular area of focus was the kinetics of binding between BSP and albumin. Many toxins and drugs are transported through the blood with the majority of the chemical bound to a circulating protein. The details of this binding process are important factors in determining the uptake rate and clearance of such chemicals by various organs. When this binding is included in kinetic models, the assumption often made is that the chemical and protein are in binding equilibrium at all times, where the concentrations of free and bound chemical are determined by the equilibrium dissociation constant (K_d). While this assumption will be true for many chemicals, a state of binding equilibrium may not exist for some chemicals (Weisiger et al., 1984; Weisiger, 1985; van der Sluijs et al., 1987; Ott and Weisiger, 1997). A lack of binding equilibrium will cause the concentration of free chemical in the plasma to differ substantially from the concentration of free chemical predicted by using the binding equilibrium assumption. Since the rate of transport of a chemical across a cellular membrane is a function of the concentration of free chemical at the membrane surface for most transport mechanisms, the degree of binding equilibrium can have a major impact on the rate at which a chemical is cleared from plasma and enters cells.

This state of non-equilibrium binding occurs when the rate at which the chemical dissociates from the circulating protein is slower than the rate at which the free chemical is used by subsequent processes such as membrane transport. Thus any free chemical in the plasma is rapidly transported into the cell, but new free chemical in the plasma accumulates slowly due to its slow release from protein. An indication that non-equilibrium binding may be occurring is data demonstrating that the membrane transport or organ uptake rate of a chemical does not correlate well with the predicted (using the binding equilibrium assumption) concentration of free chemical. Typically, for combinations of binding protein and chemical concentrations in which the equilibrium concentration of free chemical is predicted to be equal, it is found that the uptake rate is dependent on protein concentration and in fact increases for higher concentrations of protein. Chemicals for which this behavior has been seen include bromosulphophthalein (BSP), oleate, and indocyanine green (Weisiger et al., 1984; Ockner et al., 1983; and Ott and Weisiger, 1997).

Previous related studies have focused on the BSP uptake rate for exposure of the elasmobranch liver to a fixed, unvarying concentration of chemical in a single pass perfusion system (Weisiger et al., 1984) or on the 1st pass extraction fraction (Goresky, 1964; Gumucio et al., 1984; Orzes et al., 1985). The study presented here examined the effect of non-equilibrium binding between bovine serum albumin (hereafter referred to as albumin) and BSP on the uptake of BSP by the rat liver, which like other mammalian livers has a much greater BSP uptake rate than the elasmobranch liver. This greater uptake rate has the potential to shift the conditions in which non-equilibrium binding occurs. Also, the protein-binding experiments were performed in a recirculating perfused liver system with exposure to a single dose of BSP, to simulate a common toxicological kinetic situation. A previously developed biologically based kinetic model for the isolated perfused rat liver (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) was modified to include the association and dissociation rates for chemical-protein binding in the perfusate.

To this end, a series of IPRL experiments in the presence and absence of albumin has been performed. The rate of removal of the single dose of BSP from the recirculating perfusion medium for several albumin concentrations was then measured and the data interpreted according to the model. The impact of slow dissociation of chemical from protein on chemical toxicity and safety guidelines established by extrapolation from a limited set of experiments is discussed. Also, by monitoring BSP and BSP GSH concentration in perfusion medium and bile outflow, data needed to evaluate parameters in a kinetic model were obtained. The kinetic processes evaluated in this way include membrane transport through the sinusoidal membrane, bile excretion of BSP and BSP-GSH, and metabolism of BSP to BSP-GSH.

Theory

Protein-binding Kinetics

The kinetic model explicitly includes the rate of association and dissociation for the chemical-protein interaction. This enables the model to simulate situations in which the chemical and protein are not in binding equilibrium in a given compartment, such as the liver sinusoidal

compartment. Assuming a single class of binding sites on the protein, the rate at which the chemical-protein complex is formed (R_{assoc}, units μmole/s) is given by:

$$R_{assoc} = k_{on} \cdot C_{open} \cdot C_{free} \cdot V \qquad (1)$$

where k_{on} is the association rate constant ($\mu M^{-1} \cdot s^{-1}$), C_{open} is the concentration of open (unoccupied) binding sites on the protein (μM), C_{free} is the concentration of free chemical (μM), and V is the volume of the compartment in which binding is occurring (L). The rate at which the chemical dissociates from the protein (R_{dissoc}, μmole/s) is given by:

$$R_{dissoc} = k_{off} \cdot C_{bound} \cdot V \qquad (2)$$

where k_{off} is the dissociation rate constant (s^{-1}) and C_{bound} is the concentration of the chemical-protein complex (μM). The equilibrium dissociation constant K_d (μM) is then given by:

$$K_d = \frac{k_{off}}{k_{on}}. \qquad (3)$$

Liver Uptake for Protein Binding Studies

Physiologically, the uptake rate of a chemical from the sinusoidal space into the liver cells is due to the interaction of several processes at the liver membrane and within the cells. For protein binding studies, the model uses the simplifying assumption that the uptake rate is linearly proportional to the concentration of free chemical in the sinusoidal space. As long as these processes are not close to saturation, then the linear assumption will be valid. Such an assumption has been made before for perfused livers (Weisiger et al., 1984). As explored in the discussion section, the experimental conditions chosen for this study make a linear uptake process likely.

For this linear uptake, the rate at which a chemical is moved from the sinusoidal space to the intracellular space (R_{uptake}, μmoles/s) is given by:

$$R_{uptake} = k_{uptake} \cdot C_{free} \cdot V_s \qquad (4)$$

where k_{uptake} is the uptake rate constant (s^{-1}), C_{free} is the concentration of free chemical in the sinusoidal space (μM), and V_s is the volume of the sinusoidal compartment (L).

Membrane Transport for Metabolism Studies

The transport of chemical through a membrane may occur through two kinetic processes: 1) non-saturable transport which exhibits a linear dependence of transport rate on concentration; and 2) saturable transport which exhibits a hyperbolic dependence on chemical concentration, namely a Michaelis-Menten relationship.

For non-saturable transport, the rate at which a chemical on side A of a membrane is transported to side B of a membrane is given by:

$$R_{AB,nonsat} = 3.6 \cdot P_{AB} \cdot A_m \cdot C_{free,A} \qquad (\mu\text{moles/hr}) \qquad (5)$$

For saturable transport, the rate at which a chemical on side A moves to side B:

$$R_{AB,sat} = \frac{U_{AB,\max} \cdot A_m \cdot C_{free,A}}{C_{free,A} + K_{AB}} \qquad (\mu\text{moles/hr}) \qquad (6)$$

where

P_{AB}	=	diffusional permeability from compartment A to B (cm/s)
A_m	=	surface area of the membrane (cm^2)
$C_{free,A}$	=	concentration of free chemical on side A (μM)
$U_{AB,max}$	=	maximum rate of transport from side A to side B per area (umoles/hr·cm^2)
K_{AB}	=	concentration at which the transport rate is half-maximal (μM)

The factor 3.6 is to convert the units from (μmoles·cm^3)/(L·s) to μmoles/h. The model does allow P_{AB}, $U_{AB,max}$, and K_{AB} to be different than P_{BA}, $U_{BA,max}$ and K_{BA}. The total rate of transport from side A to side B is the sum of $R_{AB,nonsat}$ and $R_{AB,sat}$.

Since accurate information on membrane area is not available, the following combined parameters are reported: $PA_{AB} = P_{AB} \cdot A_m$, $UA_{AB,max} = U_{AB,max} \cdot A_m$.

Metabolism

The rate at which a parent chemical is metabolized:

$$R_{metab} = \frac{U_{\max} \cdot V \cdot C_{free}}{C_{free} + K_M} \qquad (\mu\text{moles/hr}) \qquad (7)$$

where

U_{max} = maximum rate of metabolism per volume (umoles/hr·L)

C_{free} = concentration of free chemical (μM)

K_M = concentration at which metabolism rate is half-maximal (μM)

V = volume of the compartment in which metabolism is occurring (L).

Kinetic Model

A schematic for the flows and reactions used in the model of the IPRL system is shown in Figure 1 for protein binding studies. A detailed schematic of the model used for metabolism studies is presented in Figure 2. These models were modified from previous work (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) by the addition of the association and dissociation rates of chemical from protein. A brief description of the IPRL model will be presented here.

Perfusion medium is pumped out of the medium reservoir at a constant flow rate. Free chemical, chemical-protein complex, and unbound protein then flow through the sinusoidal space of the liver. The intra-sinusoidal space and the extra-sinusoidal Space of Disse are considered to be a single compartment (called sinusoidal) for both chemical and protein (Goresky, 1980). Within the sinusoidal space, chemical and protein undergo binding reactions at the appropriate association and dissociation rates. Free chemical in the sinusoidal space is then available for uptake into the intracellular space. The subsequent disposition of chemical that is taken into the intracellular space is not modeled. The sinusoidal, intracellular, and reservoir compartments are assumed to be well-mixed. Thus, no concentration gradients exist either along the length of the sinusoid (peri-portal to peri-venous) or transverse to the flow (from cell surface to center of sinusoid).

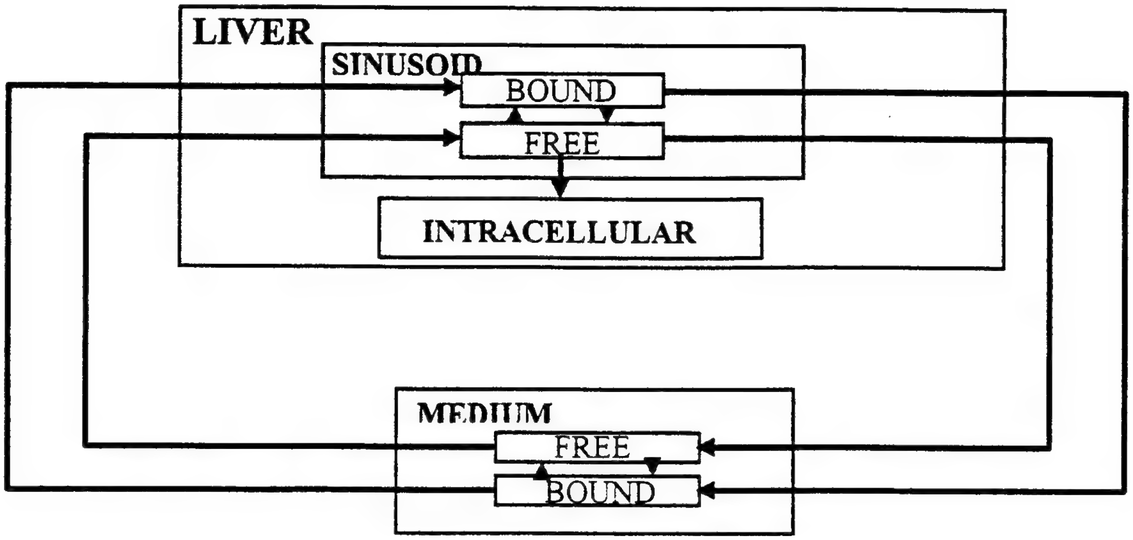


Figure 1. Schematic of the protein-binding model for the isolated perfused rat liver. The free and bound chemical recirculates through the liver via the medium reservoir. Within the sinusoidal and reservoir compartments, the free chemical can become bound and vice versa. The intracellular space takes chemical only from the free sinusoidal pool.

Methods

Liver Perfusion

Liver surgery and perfusion were performed as previously described (Wyman et al., 1995). Briefly, prior to surgery, male Fisher 344 rats weighing 200-300 g were allowed free access to food and water. After anesthetization with ether and surgically exposing the liver, the bile duct was cannulated, followed by separate cannulations of the portal vein and vena cava. The liver was excised and placed in a dish in a temperature controlled chamber. The flow rate of the perfusion medium was set to 2.4 L/h (40 ml/min) at 37°C, and the perfusion medium was oxygenated by passing it through 25 feet of Silastic tubing exposed to 95% O_2 / 5% CO_2 in a glass chamber. The pH was fixed to 7.4 by adjusting the gas flow rate as necessary. The perfusion medium was Krebs Ringer bicarbonate with 11.5 mM glucose and albumin (w/v, low endotoxin, Sigma Chem. Co. cat. no. A2934). A 30 minutes perfusion stabilization period followed surgery. A 33.5 mM taurocholate solution (Sigma Chem. Co.) was infused into the perfusion medium at 1 mL/h throughout the experiment to sustain bile production.

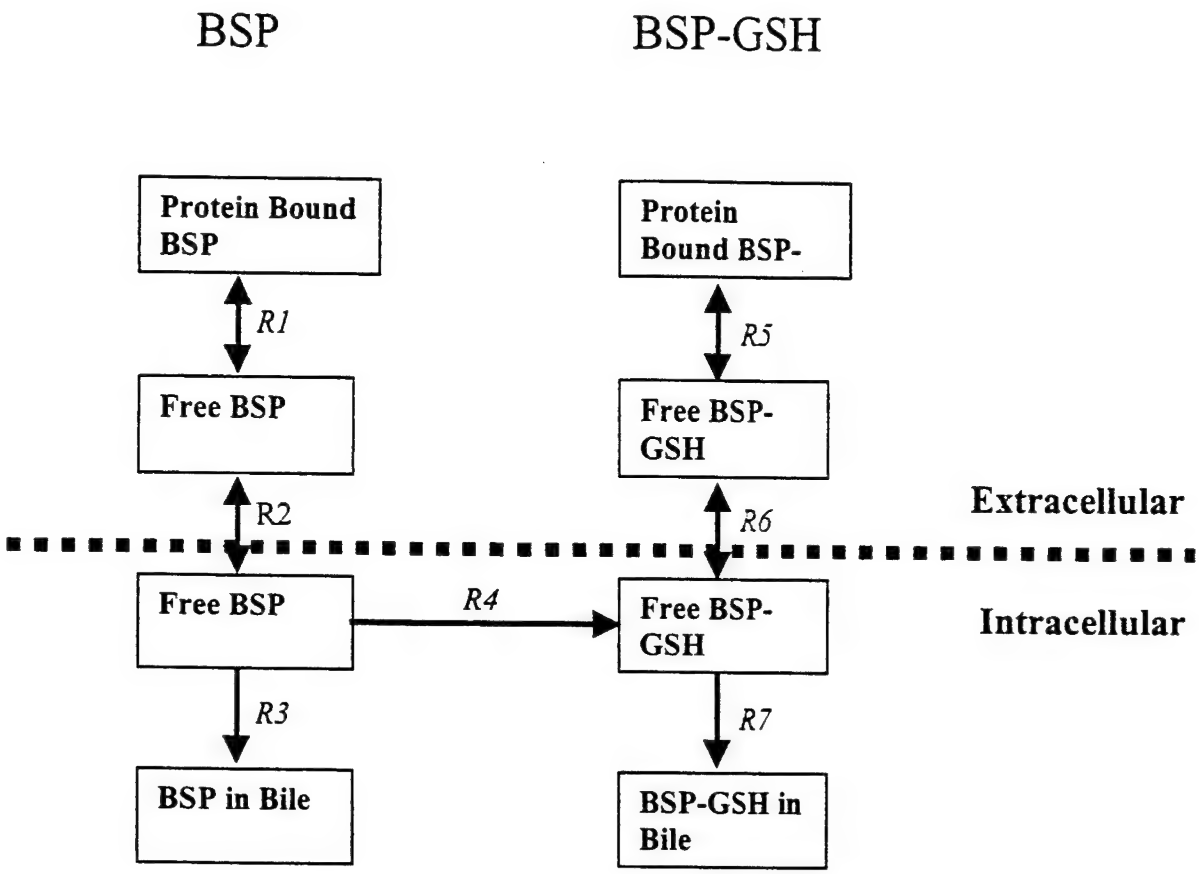


Figure 2: A schematic for the liver-specific processes in the kinetic model used for analysis of metabolism experiments. The compartments on the left and right sides refer to BSP and BSP-GSH, respectively. Each rate R (μmoles/hr) may represent a linear and/or a Michaelis-Menten kinetic process. The entire model also includes flow between a reservoir and the extracellular space as appropriate for a recirculating or one-pass experiment. Note that the extracellular compartment is assumed to include sinusoids and all other extracellular volume.

For recirculating perfusions, the following protocol was used. Four μmoles of BSP were added to 200 mL of perfusion medium for an initial BSP concentration of 20 μM. The duration of perfusion was 150 min. BSA concentrations were 0% (n=3), 0.25% (n=3), 1% (n=3), and 4% (n=4) (w/v; concentrations = 0, 37, 143, 570 μM).

For one-pass perfusions, the following procedures were used. The duration of perfusion with BSP was 60 min. Four BSP:BSA combinations were used: (1) 10 μM BSP and 0% BSA (n = 2), (2) 40 μM BSP and 0 μM BSA (n = 2), (3) 20 μM BSP and 0.25% BSA (n = 3), and (4) 20 μM BSP and 1% BSA (n=1)

At several time points following addition of BSP, a 0.5 ml aliquot of perfusion medium was removed from the medium reservoir. The bile outflow was collected over 30 minute time intervals, and bile samples were stored at 0°C until analysis. The concentration of total BSP (BSP plus metabolites such as BSP-glutathione) in perfusion medium was determined by mixing the medium sample with an equal part (0.5 ml) of 1M NaOH and measuring spectral absorbance at 580 nm. For measurement of total BSP in bile, 10 μL of bile sample was added to 1 mL perfusion medium, which was then mixed with 1 mL of 1M NaOH. The spectral absorbance was again measured at 580 nm. Relative BSP and BSP-GSH amounts were quantified in perfusion medium and bile using HPLC.

Model Implementation

The biologically based kinetic model was coded on a PC using Advanced Computing Simulation Language (ACSL level 11, MGA Associates, Concord, MA), a numerical integration package.

Model Parameters

The volumes (ml) of the sinusoidal and intracellular compartments are assumed to be 21.7% and 58.3%, respectively, of the weight of the liver (g) (Goresky, 1980).

Regarding binding of BSP to albumin, three studies have found multiple high-affinity binding sites on albumin with K_d less than 0.26 μM. In one study, one binding site had a K_d of 0.06 μM,

and two more binding sites with K_d of 0.63 μM (Baker and Bradley, 1966), while a second study found two to three binding sites with a K_d of 0.26 μM (Pfaff et al., 1975). Another study also found a high affinity binding site (K_d = 0.05 μM) but found less than one binding site per albumin molecule at that affinity (Zhao et al., 1993). Since the first two studies both estimate 3 strong binding sites, they are used as a basis for choosing binding parameters. Accordingly, a K_d of 0.2 μM with 3 binding sites per albumin is chosen. Additional lower affinity binding sites on albumin will have a negligible effect on the kinetics due to the low concentration of BSP relative to albumin in all experiments.

Fitting and Error Analysis

These considerations leave two independent parameters to be determined by fitting the reservoir BSP concentration versus perfusion time data: k_{off} and k_{uptake}. The critical information needed to determine these parameters is the relative uptake rates of BSP at different albumin concentrations. Therefore, the data set used for parameter estimation consisted of data from all three albumin concentrations. Application of a non-linear least squares fitting procedure then produced the best estimates of the parameters. In addition, for one fit, k_{off} was fixed to a very large value to explore the kinetics when binding equilibrium is established in the sinusoidal space. Note that setting k_{off} to a large value forces k_{on} to be large due to Eq. 3.

Experimental values are presented as mean ± standard deviation. For fit parameters, 90% confidence intervals were determined by finding the range of each parameter that kept chi-squared within 2.71 of its minimum value (Press et.al., 1992). In other words, the best fit occurs when chi-squared (residuals divided by standard deviations. squared, and summed for each data point) is at its minimum value. The parameter of interest is then fixed to a value that is not its best fit, and the non-linear least squares algorithm is applied again with all other parameters variable. Chi-squared will be larger, due to the less ideal fit. The 90% confidence interval occurs when a chosen value for the parameter of interest causes chi-squared to increase by 2.71. These confidence intervals are presented in parentheses after the best fit value.

Results

Protein Binding Studies

Three livers were perfused at each albumin concentration. The average wet liver weights for the perfusions were (in g) 8.63 ± 0.34, 10.39 ± 0.78, and 10.76 ± 1.21, for the 4%, 1%, and 0.25 % albumin concentrations respectively. The volumes of each liver compartment in the model were scaled according to the average weight for each albumin concentration.

BSP concentration in the perfusion medium versus time for each of the 3 albumin concentrations is shown in Figure 3A. In the presence of 4% albumin the concentration of BSP declines the most slowly, which is expected due to the low concentration of free BSP under this condition. The lines in Figs. 3A and 3B are from the modeling discussed below.

The results of the modeling are presented as lines in Fig. 3A. The 2 fit parameters are a k_{off} value of 0.114 (0.097 to 0.133) s^{-1}, and a k_{uptake} of 156 (139 to 175) s^{-1}. This k_{off} value, when combined with the equilibrium dissociation constant K_d according to Eq. 3, produces an estimate of the association constant k_{on} of 0.569 (0.486 to 0.667) s^{-1} μM^{-1}. Note that the large value of k_{uptake} relative to k_{off} does not imply that the uptake flux is greater than the dissociation flux since the uptake flux is proportional to k_{uptake} times the extremely low concentrations of free BSP (Eq. 4), while the dissociation flux is proportional to k_{off} times the substantially larger concentration of bound BSP (Eq. 2).

The data was also fit with the assumption that k_{off} equals 2.78 s^{-1} (10,000 hr^{-1}), a large value relative to the 0.114 s^{-1} fit above. This assumption will tend to produce an equilibrium binding situation in the sinusoids. Fig 3B presents the same experimental data as Fig 3A, but the lines represent the predicted reservoir concentration with the equilibrium binding assumption. In this

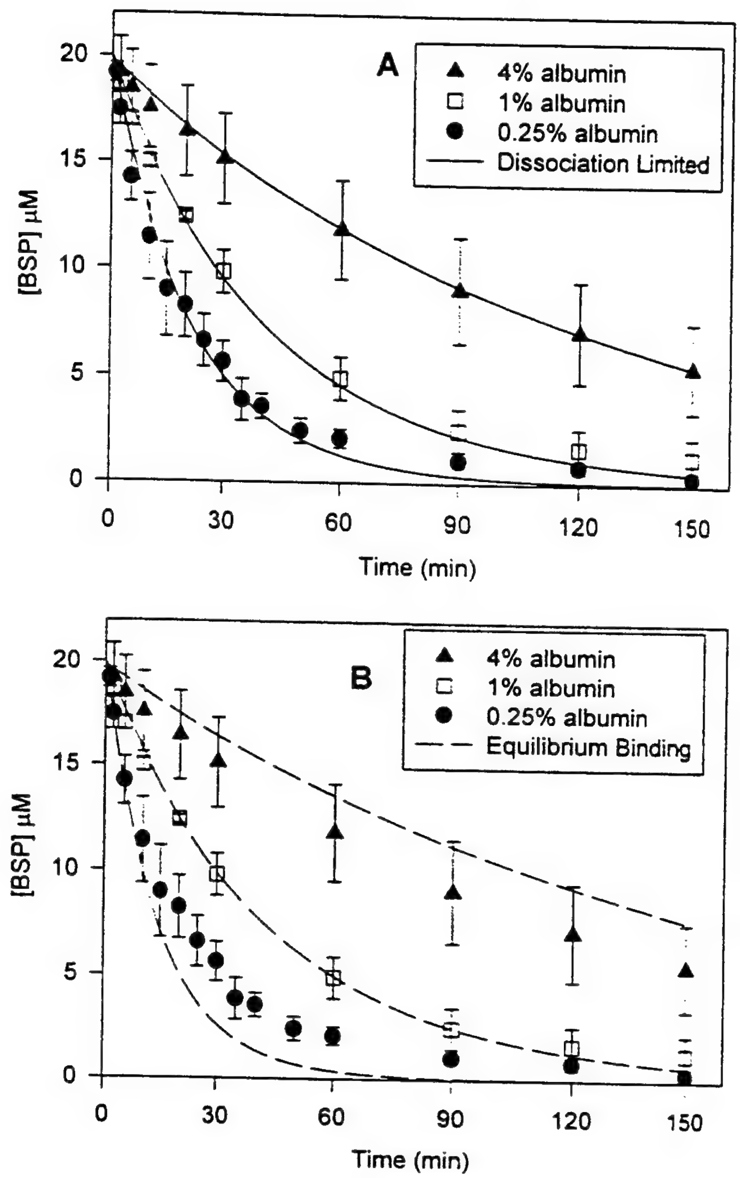


Figure 3. Reservoir concentration of BSP as a function of perfusion time. Symbols represent experimental measurements of BSP concentration, and error bars are standard deviation. **(A)** Dissociation limited case: lines represent best fit prediction from model, with k_{off} estimated to be 0.114 s^{-1} and k_{uptake} estimated to be 156 s^{-1}. **(B)** Equilibrium binding case: symbols represent same data set as in (A). Lines represent prediction from model with k_{off} fixed to 2.78 s^{-1}. k_{uptake} is estimated to be 89 s^{-1}.

case, the data set was fit by varying k_{uptake} alone. The k_{uptake} value for this fit was 89 (81 to 106) s^{-1}. This best fit accurately predicted the 1% albumin data, but was substantially in error for the 4% and 0.25% cases.

To illustrate the prediction of non-equilibrium binding in the sinusoids, Table 1 presents estimates from the model of the free concentration of BSP in the sinusoids when the total BSP in the sinusoid is predicted to be 10 μM. This 10 μM total sinusoidal BSP concentration will occur at a different time for each of the protein concentrations and fitting protocols. For each of the fitting procedures, corresponding to the fits in Figs 3A and 3B, the free concentration of BSP as predicted by the model is compared to the free concentration of BSP that would be present if all the BSP in the sinusoidal space were in equilibrium with albumin. For the dissociation limited fit corresponding to Fig 3A, the predicted free concentration of BSP for the two lower albumin concentrations is substantially different (lower) than the value that would be expected if binding equilibrium existed. Under the fast binding scenario in which k_{off} was set to a high value, BSP is near binding equilibrium for all protein concentrations as expected.

Table 1 – Simulated sinusoidal free BSP concentrations when total BSP in the sinusoid is 10 μM.

	Albumin	0.25%	1%	4%
Dissociation Limited Fit (k_{off} fit to 0.114 s^{-1})	Total BSP (μM)	10.0	10.0	10.0
	Free BSP, Model Prediction (nM)	5.6	2.9	1.0
	Free BSP, Binding Equilibrium (nM)[a]	20.6	4.8	1.2
Fast Binding Fit (k_{off} fixed at 2.78 s^{-1})	Total BSP (μM)	10.0	10.0	10.0
	Free BSP, Model Prediction (nM)	20.3	4.8	1.2
	Free BSP, Binding Equilibrium (nM)[a]	20.6	4.8	1.2

[a]Calculated from albumin and total BSP concentrations assuming 3 binding sites per albumin and a K_d of 0.2 μM.

Another way to distinguish between the purely dissociation limited condition and conditions which may be limited partly by dissociation and partly by intrinsic liver uptake is to compare the relative value of the rate of uptake (R_{uptake}) with the rate of dissociation (R_{dissoc}). From Table 1, the model prediction values for the dissociation limited fit can be used to provide estimates for these rates when total sinusoidal BSP is 10 μM. The values are calculated according to Eqs. (2) and (4), normalized to liver weight for each protein concentration, and presented in Table 2. Note that for 0.25% albumin, R_{uptake} is almost as great as R_{dissoc}. Since the rate at which *free* BSP enters the liver via perfusion is quite small for all albumin concentrations (less than 1% of R_{uptake} per g liver for each data point in Table 2), the majority of free BSP for uptake must come from dissociation. Therefore, R_{uptake} can not exceed R_{dissoc} for any albumin concentration, and the similarity of these values for the 0.25% case indicates that the dissociation rate is the primary factor in determining uptake rate. For 1% albumin, the dissociation rate exceeds the uptake rate by a greater amount, indicating that dissociation is not as limiting in this case. However, the lack of binding equilibrium as shown in Table 1 for the 1% case indicates that dissociation limitations do affect the net uptake rate. Thus the 1% albumin concentration experiment appears to produce an intermediate regime where both the intrinsic uptake rate and the dissociation rate control the net uptake rate. For 4% albumin, the large value of R_{dissoc} relative to R_{uptake} indicates an even smaller importance of dissociation rate in determining net uptake rate. Except for a very small correction due to differing inflow and outflow of free BSP, mass balance considerations force R_{assoc} to equal the difference between R_{dissoc} and R_{uptake}. Thus, in the 4% case, R_{assoc} is close in value to R_{dissoc} since so little of the dissociated chemical is taken up by the cells.

Table 2 – Model predictions for uptake rate and dissociation rate, normalized to average liver weight for each protein concentration. Values in μmoles/(min·g liver).

	0.25%	1%	4%
R_{uptake}/g liver	1.13×10^{-2}	0.58×10^{-2}	0.20×10^{-2}
R_{dissoc}/g liver	1.48×10^{-2}	1.49×10^{-2}	1.50×10^{-2}

Metabolism Studies

Liver Viability

For recirculating experiments, LDH enzyme leakage began to appear after 60 min of exposure to BSP. Enzyme leakage, expressed as the percentage of total LDH originally in liver which appears in the reservoir after 2.5 hr of BSP exposure, was 14.2 ± 2.9%, 12.7 ± 7.5%, and 6.8 ± 4.1% for the BSA concentrations of 4%, 1%, and 0%, respectively. For recirculating experiments, the bile flow reached peaks of 1.37 ± 0.17, 0.87 ± 0.05, and 0.86 ± 0.20 $\mu L \cdot min^{-1} \cdot g$ $liver^{-1}$ during the second 30 min collection interval for the 4%, 1%, and 0% BSA concentrations, respectively. The bile flow gradually dropped off to 60 to 80% of peak values by the end of the 2.5 hr BSP exposure. For the one-pass experiments, the bile flow averaged 0.81 ± 0.30, 0.78 ± 0.29, 1.15 ± 0.14, and 0.99 ± 0.17 $\mu L \cdot min^{-1} \cdot g$ $liver^{-1}$ for the BSP (µM):BSA (%) ratios of 10:0, 40:0, 20:1, and 20:0.25, respectively. The bile flow for the second 30 min period was more than 70% of peak bile flow for all livers.

Recirculating Experiments

At lower BSA concentrations, the rate of BSP uptake is much higher (Fig. 4). This is expected due to the increased concentration of free BSP in the extracellular compartments in the absence of protein binding. With no BSA, essentially all BSP is excreted into bile in 2.5 hrs (Fig. 5A). At 4% BSA, due to the slower uptake rate, just over half of the BSP is excreted into bile. The percent of total BSP excreted as BSP-GSH in the bile was higher for 0% BSA than for 4% BSA (Fig. 5B). At first glance, this is surprising since the 0% case has a greater rate of BSP uptake, and therefore needs a *much* greater rate of metabolism (than is occurring in 4% case) to produce this higher conjugation fraction

One-pass Experiments

The outflow BSP + BSP-GSH concentration was highest when inflow concentration was 40 µM with no BSA (Fig 6A). The outflow concentration was lowest when inflow concentration was 10 µM and 0% BSA. When BSA was present, the total outflow concentration was changed very little from the inflow concentration (20 µM). The higher the BSA concentration, the less change in outflow compared to inflow. The percentage of total BSP appearing as BSP-GSH in the

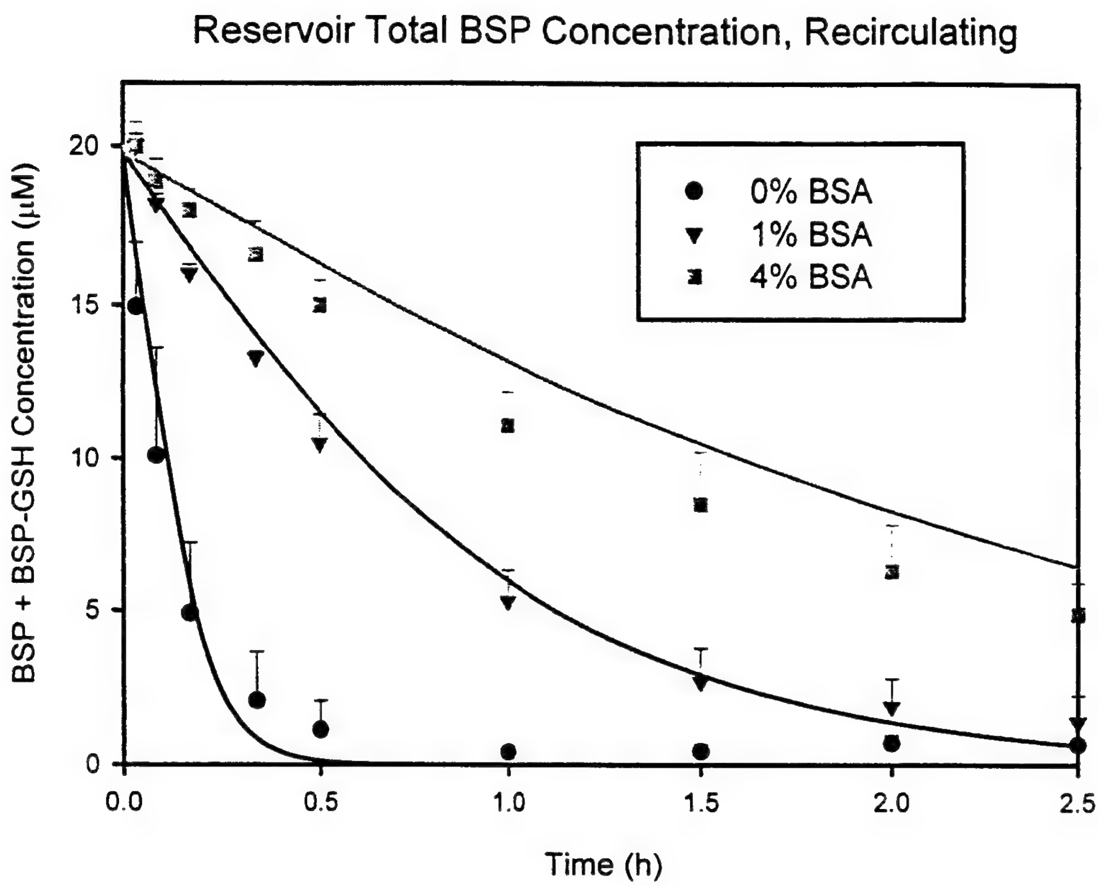
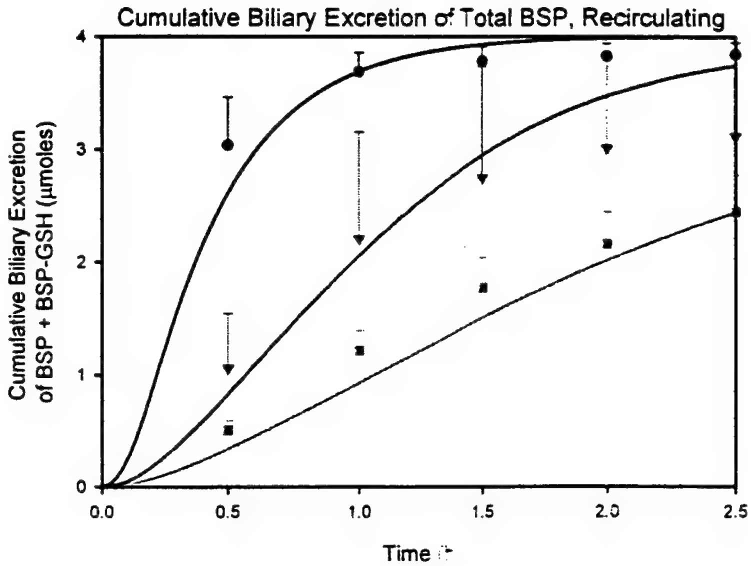


Figure 4: Total BSP (BSP + BSP-GSH) concentration in perfusion medium vs. time of exposure to BSP for recirculating perfusion experiments. Symbols represent experimental results (expressed as mean ± SD), and lines represent the model predictions (Table 3). The initial BSP concentration was 20 μM, and the BSA concentration was 0, 1, or 4% (w/v).



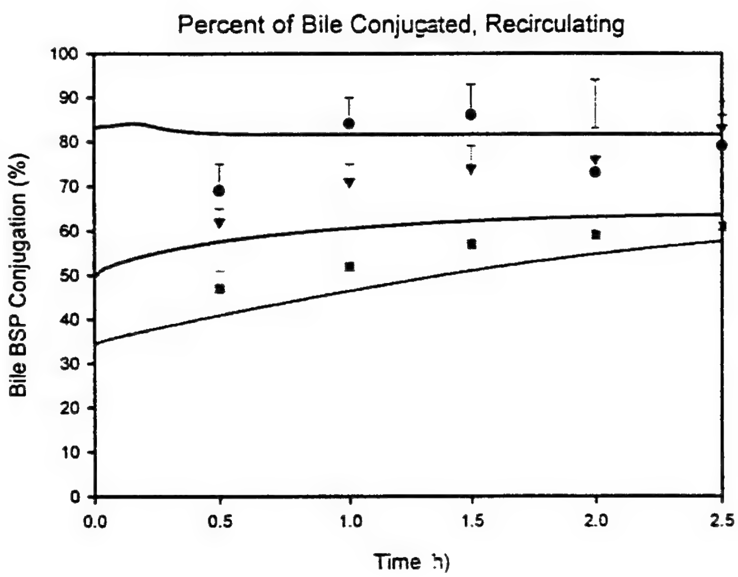


Figure 5: BSP in bile outflow for recirculating perfusion experiments. Symbols represent experimental results (expressed as mean ± SD), and lines represent the model predictions (Table 3). **(A)** Cumulative level of BSP + BSP-GSH in bile vs. time of exposure to BSP. **(B)** Percentage of total BSP in the bile that was present as BSP-GSH vs. time of exposure to BSP.

perfusion medium outflow was highest when BSA was absent and 10 μM BSP was used (Fig 6B). With 40 μM BSP, the conjugated percentage went down substantially. When BSA was present, no conjugated BSP was seen in medium outflow. The two experiments without BSA produced the highest BSP + BSP-GSH bile output (Fig 7A). At 30 min, these two had similar total bile output. During the second 30 min period, the 40 μM BSP, 0% BSA experiment had a reduced BSP + BSP-GSH output into the bile. This may indicate a significant decline in liver function for this condition, although the rate of total BSP uptake for this condition remained steady (Fig 6A). When BSA was present, the total bile output was lower, with the smallest total BSP output occurring for the 1% BSA condition. In the absence of BSA, over 80% of total BSP in bile was present as BSP-GSH. With BSA, the fraction dropped to 64% (0.25% BSA) or 51% (1% BSA).

Modeling

Table 3 lists the modeling parameters that were used to produce the lines in Figures 4 to 7. The method for choosing these parameters is described below.

Table 3: Parameters Used in Model

R1	N = 3 binding sites	R5	N = 3 binding sites
	$k_{off} = 410\ h^{-1}$		$k_{off} = 41000\ h^{-1}$
	$k_{on} = 2050\ (\mu M \cdot h)^{-1}$		$k_{on} = 2050\ (\mu M \cdot h)^{-1}$
R2	PA = 1.44 l/h (bi-direct.)	R6	PA = 1.44 l/h (bi-direct.)
	UA_{max} = 60 μmoles/h (bi-direct.)		UA_{max} = 60 μmoles/h (bi-direct.)
	K = 0.05 μM		K = 0.2 μM
R3	PA = 0.0 l/h	R7	PA = 0.0 l/h
	UA_{max} = 3 μmoles/h (one way)		UA_{max} = 15 μmoles/h (one way)
	K = 0.002 μM		K = 0.1 μM
R4	$U_{max} \cdot V$ = 30 μmoles/h		
	K_m = 0.02 μM		

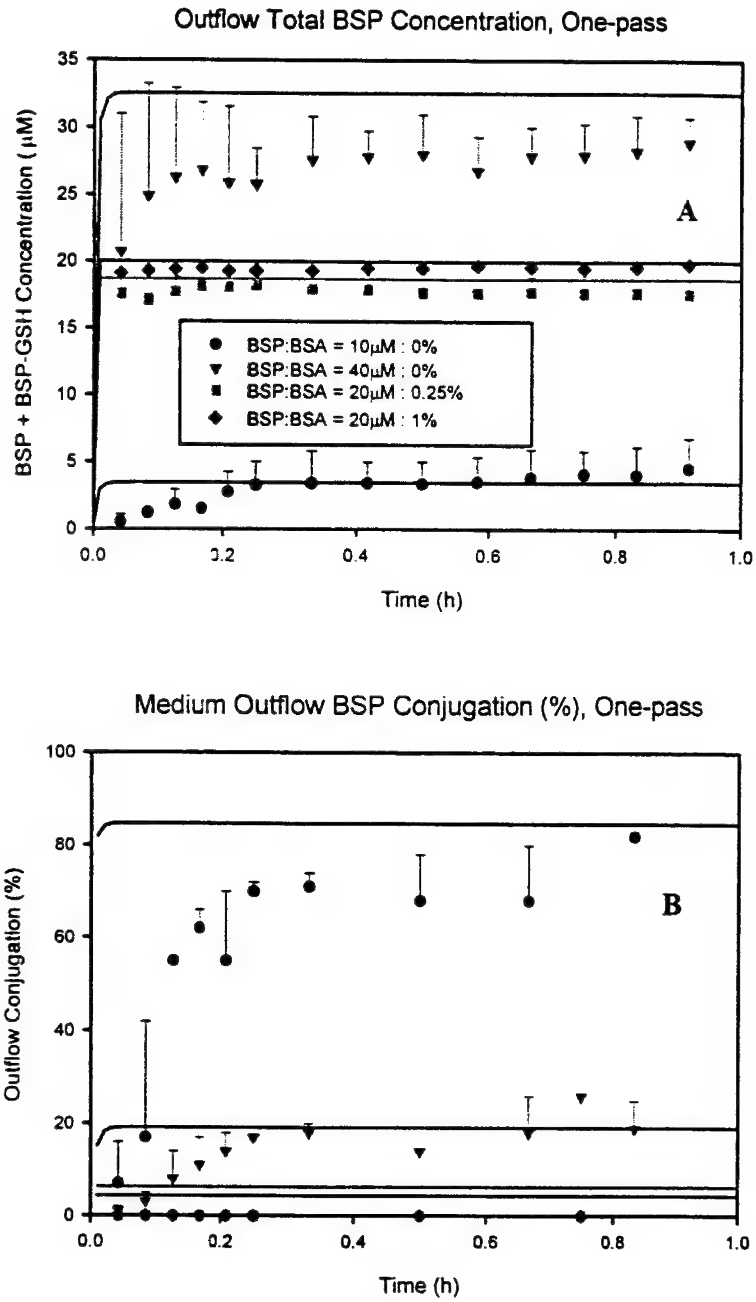
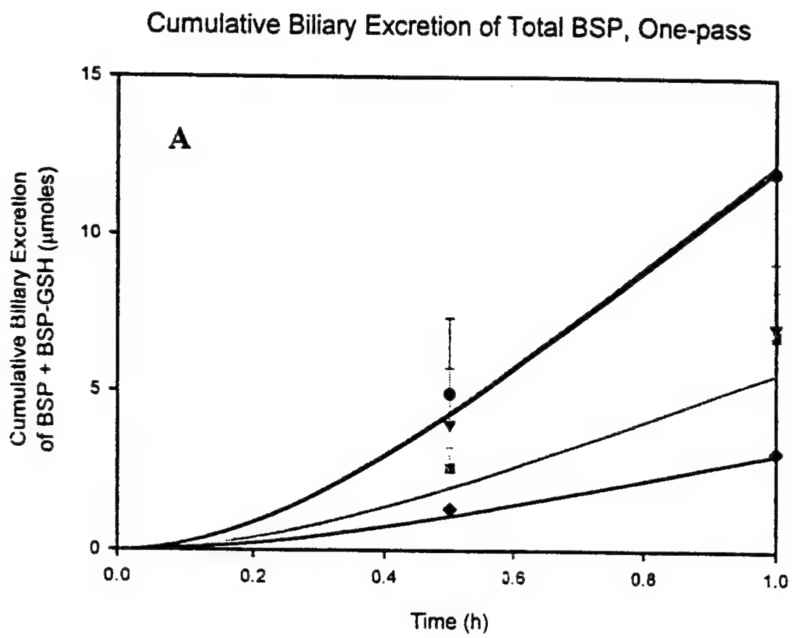


Figure 6: BSP in perfusion medium outflow for one-pass perfusion experiments. Symbols represent experimental results (expressed as mean ± SD), and lines represent the model predictions (Table 3). **(A)** Total BSP (BSP + BSP-GSH) concentration in the medium outflow vs. time of exposure to BSP. **(B)** Percentage of total BSP in the medium outflow that was present as BSP-GSH vs. time of exposure to BSP.



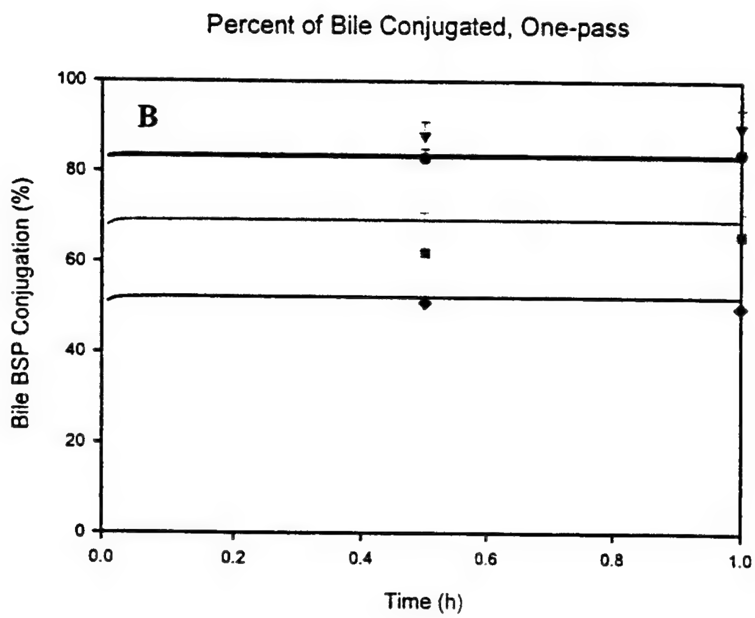


Figure 7: BSP in bile outflow for one-pass perfusion experiments. Symbols represent experimental results (expressed as mean ± SD), and lines represent the model predictions (Table 3). **(A)** Cumulative level of BSP + BSP-GSH in bile vs. time of exposure to BSP. **(B)** Percentage of total BSP in the bile that was present as BSP-GSH vs. time of exposure to BSP.

Protein Binding

R1: The parameters for binding of BSP to BSA are the same as was used in a previous study (Foy *et al.*, 1999). These parameters were based on earlier studies (Baker and Bradley, 1966).

R5: Previous reports indicate that the dissociation constant K_d (= k_{off}/k_{on}) of BSP-GSH is 2 orders of magnitude larger than for BSP (Baker and Bradley, 1966; Geng *et al.*, 1995). Therefore, k_{off} was set to a value 100 times larger than for BSP. Multiple binding sites have also been detected, so 3 binding sites per BSA molecule were again used.

Sinusoidal Membrane Transport

R2: A previous study performed only in presence of BSA fixed the ratio of UA_{max} to K (Foy, 1999). Since BSA was present in this earlier study, the free extracellular BSP concentrations were quite low and the UA_{max} value was not determined. The UA_{max} value in Table 1 was based on a published study of the high-affinity BSP transporter in cultured hepatocytes (Sorrentino *et al*, 1994). However, the experimental conditions of the current study also included cases without BSA, which leads to much higher free BSP concentrations. The single high affinity transport system described by UA_{max} and K was not able to accurately predict the experimental data, and so a second non-saturable transport parameter PA was introduced. Its value was iteratively determined. The presence of a second low affinity transport system for BSP transport has been documented before (Sorrentino *et al*, 1994; Schwenk *et al*, 1976).

R6: Previous studies have determined a maximum uptake rate for BSP-GSH of 50 μmole/hr (for a 10 g liver) (Geng *et al*, 1995). Since this is so close to the 60 μmole/hr value used for BSP uptake (R2), the decision was made to make both UA_{max} values equal to 60 μmole/hr. No information was found for low affinity transport, so PA was also kept the same as for R2. K was set as described below.

Bile Membrane Transport

R3: When no BSA was present in the 1-pass experiment, the bile production of non-conjugated BSP reached a maximum of around 2.0 μmole/hr (20% of 10 μmoles in 1 hour) for both 10 μM and 40 μM BSP. The UA_{max} was fixed to a value slightly higher than this to account for

time lag in bile excretion. The finding in recirculating experiments that more BSP was conjugated in absence of BSA also supports the probability of saturated bile output of non-conjugated BSP. The K value here had to be smaller than the K for metabolism (R4) in order to produce the lower conjugation fractions when BSA was present. If the Ks were equal or reversed, then BSP would always prefer metabolism (R4) to excretion (R3) and predicted conjugation fractions would be higher than experimental. No PA value was found to be necessary. The transport was one-way, from intracellular space to bile. This is consistent with the highly elevated concentrations of BSP in bile.

R7: The total conjugated BSP bile output in absence of BSA was ~ 10 μmoles in 1 hour for both 10 μM and 40 μM experiments. Thus UA_{max} was set to 15 μmoles/h (again, has to be slightly higher due to lag). The K was set relative to K for R6 so that little BSP-GSH escaped to medium when BSA was present.

Metabolism

R4: The total BSP conjugated in 1 hour in the absence of BSA for the 10 μM experiment was 18 μmoles, and for 40 μM experiment was 15 μmoles. Thus the V_{max} was set slightly higher than this value. The K_m value was set relative to K for R3 so as to achieve the proper conjugation fractions when BSA was present.

Thus the UA_{max} for R3 and R7, and V_{max} for R4, have been determined, based primarily on the fact that these rates appear to be saturated for the one-pass experiments in the absence of BSA. The K values are generally only fixed relative to each other. K for R3 must be less than K for R4, and K for R7 must be less than K for R6. Determining absolute values for K will require a careful analysis of the intracellular BSP concentrations and the degree of intracellular binding.

Discussion

A biologically-based kinetic model which predicts the BSP and BSP-GSH kinetic profiles in perfusion medium and bile under a variety of experimental conditions has been developed. The model predicts complex experimental findings, such as the increased percentage of conjugated BSP in the bile when the liver BSP uptake rate is higher in the absence of circulating BSA. The

integration of widely varying experimental conditions, namely one-pass perfusion vs. recirculating perfusions and presence vs. absence of BSA, has proved useful in quantifying key parameters of the system. This detailed model of BSP kinetics can serve as a tool to identify and quantify toxic effects of other chemicals in the perfused liver conditions.

The experiments and modeling presented here also indicate that the rate of dissociation of a toxin from a protein may play an important role in determining the ultimate kinetics for the toxin. Using the assumption that the dissociation of BSP from albumin was so rapid that equilibrium was maintained in the sinusoids produced an inaccurate prediction of experimental liver uptake kinetics. On the other hand, allowing non-equilibrium BSP binding in the sinusoids enabled the model to match experimental data for 0.25, 1, and 4% albumin while using literature values for the equilibrium dissociation constant.

The ability to extrapolate from a few experimental test cases to a range of possible exposure doses and conditions is a major goal of predictive toxicology. Errors in the underlying extrapolation model can lead to inaccurate predictions. For example, errors in making the assumption of equilibrium binding in the sinusoidal spaces can be seen in Fig. 3B. If the time required for BSP concentration to fall from 20 μM to 3 μM were deemed critical, due possibly to BSP toxicity at a non-liver organ, then the equilibrium-binding prediction of 22 min in the presence of 0.25% albumin is considerably shorter than the experimental finding of 38 min by nearly a factor of two. This may lead to safety exposure limits that are overly risky. The converse problem, predicting clearance by the liver that is slower than reality, will occur if one uses the fast binding parameters at 4% albumin. Although albumin is present in rat plasma at approximately a 4% concentration, under *in vivo* conditions the number of free BSP (or other chemical) binding sites on albumin is likely to be quite variable, due to competing chemicals in the plasma and variable plasma albumin concentrations (under pathologic liver conditions). Thus any predictive model will need to explore a range of chemical:albumin ratios when attempting to model the *in vivo* kinetics of a compound.

Although little data exist regarding the dissociation and association rate constants for the majority of toxic compounds, one can expect that other toxic compounds, especially other

anionic compounds, may be so tightly bound to albumin that they behave in a fashion similar to BSP. Using low K_d values as an indication of strong binding and possibly slow dissociation, the phenoxyacetic acid class of pesticides has been shown to bind to bovine serum albumin with K_d values as low as 0.4 μM, which is similar to the K_d for BSP of 0.2 μM (Fang and Lindstrom, 1980). Also, the insecticide chlorpyrifos has been found to bind to albumin with a K_d of 3.4 μM (Sultatos et al, 1984). Plasma carriers other than albumin may also contribute to a potential dissociation limited condition. For example, alpha 1-acid glycoprotein has been shown to bind the antibiotics lincomycin and clindamycin with K_d values ranging from 1 to 3 μM (Son, et al., 1998).

Some sense of when dissociation limited conditions are likely to affect a toxicokinetic analysis can be gained by identifying such regions on a graph of total concentration of protein binding sites vs. total chemical concentration. At one extreme, dissociation limitations will not occur when the majority of the chemical is free, as opposed to being bound to the protein. When most of the chemical is free, the effect of slow dissociation of the remaining small fraction of bound chemical will be small. The transition zone for the situation in which more than 90% of the chemical is free occurs when:

$$k_{on} \cdot C_{open} < \frac{k_{off}}{9}. \tag{8}$$

At the other extreme, dissociation limited conditions will no longer dominate when binding equilibrium in the sinusoid occurs. Binding equilibrium will occur when the rate of uptake by the liver is smaller than the rates at which chemical associates with or dissociates from the protein. Binding equilibrium will occur at higher concentrations of protein since high concentrations of protein produce low uptake rates (due to low concentrations of free chemical) and high rates of binding. Thus, the transition to binding equilibrium will occur when the rate at which chemical associates with protein is greater than the rate at which chemical is taken up by the liver, or

$$k_{on} \cdot C_{open} > k_{uptake}. \tag{9}$$

After converting the inequalities in Eqs. (8) and (9) to equalities, these boundary lines are graphed in Fig. 8 for a range of albumin binding site concentrations. The rate constants and

values for C_{open} used to create the boundary lines in the graph are those determined from the fits and the BSP-albumin binding parameters presented above. These lines are meant to indicate roughly where transitions occur from one behavior to another. When an experiment produces concentrations near one of these boundary lines, the behavior will actually be intermediate in nature.

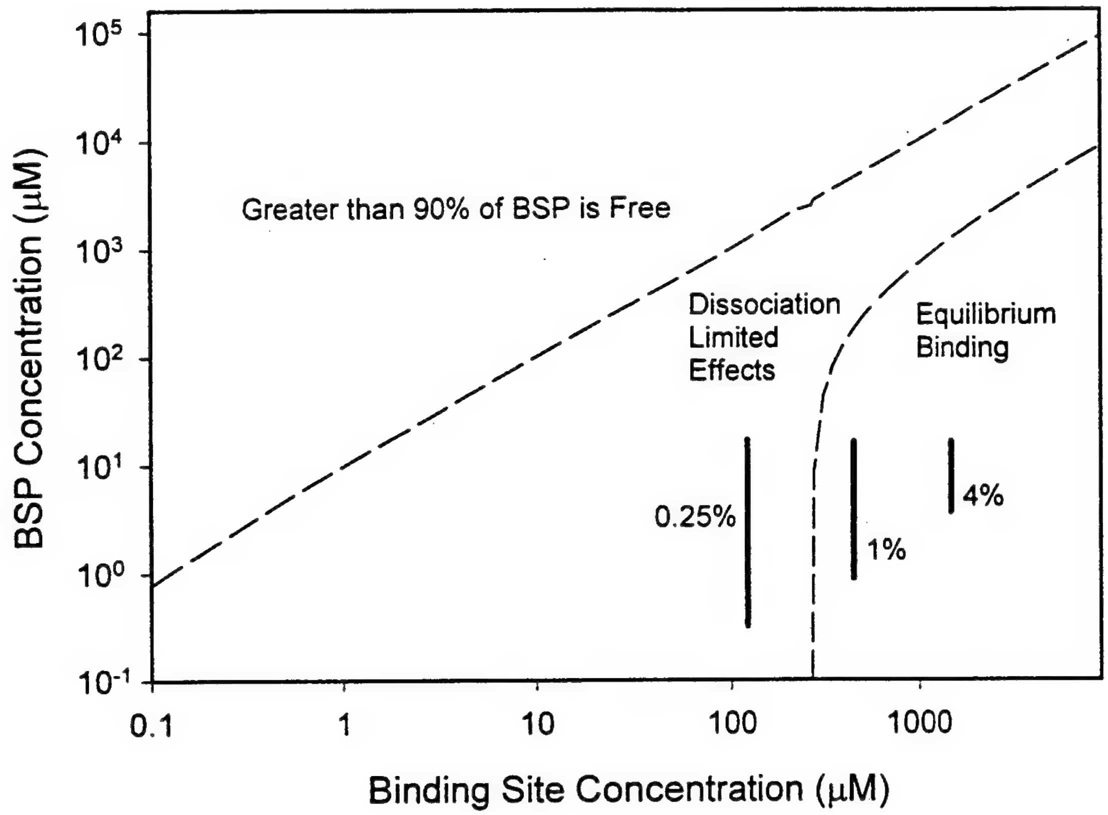


Figure 8. Identification of the chemical and protein concentrations that tend to produce dissociation limited conditions. The dashed lines are mark the boundaries between regions. Solid vertical lines mark the conditions that occurred in experiments in this study. The parameters used to generate these curves are $k_{off} = 0.114\ s^{-1}$, $k_{on} = 0.569\ s^{-1} \cdot \mu M^{-1}$, $k_{uptake} = 156\ s^{-1}$.

One can see that at lower BSP and higher binding site concentrations, the likelihood of dissociation limited conditions is less due to the establishment of near binding equilibrium. Similarly, at high BSP and low binding site concentrations, most BSP will be free and the net uptake of BSP by the liver will be limited by the rate of flow into the sinusoid (perfusion limited). The location of the experiments in this study are also marked on the graph, indicating that the 4% and 0.25% experiments are solidly on either side of the binding equilibrium boundary, while the 1% experiment is near the boundary line (and thus in a transition zone) between dissociation limited and equilibrium binding conditions. The location of the boundary lines will be different for different chemicals, proteins, and organs due to different parameters in Eqs. (8) and (9). In fact, many chemicals may not exhibit dissociation limited effects for any combination of binding site and chemical concentrations. Dissociation limited effects will be minimal when intrinsic uptake by the liver is slow relative to the rate at which chemical dissociates from the protein binding site. Graphically, this corresponds to a chemical/protein interaction in which the right-hand boundary line (defined by Eq. (9)) has moved to the left, and/or the left-hand boundary line (defined by Eq.(8)) has moved right such that the dissociation limited region no longer exists. For such a chemical/protein interaction, a single transition in the rate-limiting process will occur as the binding site concentration increases, and this transition would be defined not by Eqs. (8) or (9), but rather by the transition from flow-limited conditions to uptake limited conditions (see Weisiger, 1984, for details).

Fig. 9 presents another way to view the dissociation rate limitation on liver uptake. Here, for a given concentration of BSP in the reservoir (10 μM) the rate of uptake into the cells is plotted for a range of concentrations of protein binding sites. Two scenarios are plotted: one which uses the same parameters determined by fitting the experiments in this paper, and one which uses all the same parameters except k_{off} is increased from 0.114 s^{-1} to 2.78 s^{-1} and k_{on} is increased from 0.569 $s^{-1} \cdot \mu M^{-1}$ to 13.9 $s^{-1} \cdot \mu M^{-1}$. These changes keep K_d at 0.2 μM, but setting k_{off} to a high value places the second scenario into a binding equilibrium condition in the sinusoid. The general trend, that uptake rate decreases for increasing binding site concentration, occurs in both scenarios. However, the lower k_{off} in the first case produces a lower uptake rate for a range of binding site concentrations between 1 and 250 μM. The two curves approach each other at low

binding site concentrations because both are being limited by the rate at which new chemical is flowing into the liver. At high binding site concentrations, the uptake rate becomes very small due to the small concentration of free chemical, and this low uptake rate enables the chemical and protein to approach binding equilibrium even when k_{off} is small. The data presented in Table 1 also illustrates the phenomenon that higher protein concentrations tend to promote the binding equilibrium condition (Weisiger, 1985; Sorrentino et al., 1994).

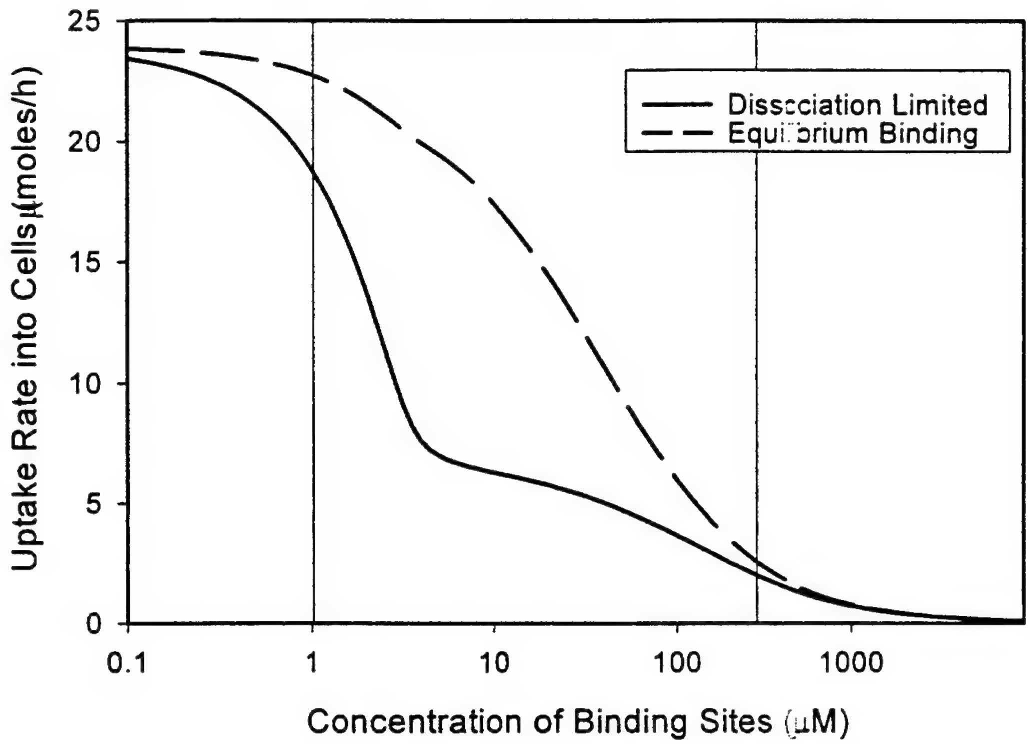


Figure 9. Simulated uptake rate into the intracellular space for a range of binding site concentrations. For these simulations, the initial reservoir BSP concentration was 20 μM and R_i was evaluated at the time points when the reservoir concentration dropped to 10 μM. For the dissociation limited curve, in addition to the parameters listed in Fig. 5, flow Q was 2.4 L/hr and the liver weight was 10g. For the equilibrium binding curve, model parameters were identical except $k_{off} = 2.78\ s^{-1}$ and $k_{on} = 13.9\ s^{-1}\cdot\mu M^{-1}$. The vertical lines correspond to boundary lines identified in Fig. 3 at a BSP concentration of 10 μM.

Regarding the dissociation rate constant determined in this work, no previous direct measurement of k_{off} appears to have been made for BSP-albumin binding. An estimate of k_{off} between 0.053 and 0.208 s^{-1} has been made from application of a model to perfused liver data (Weisiger et al., 1984). The value of 0.114 s^{-1} found in this work is within this range. Measurement of k_{off} for a related compound, dibromosulfophthalein (DBSP), revealed a value of 0.047 s^{-1} (van der Sluijs et al., 1987). Smaller values for k_{off} would tend to promote the dissociation limited condition.

Uptake of BSP by the liver has been treated in a simplified linear manner, with net uptake rate being proportional to the concentration of free BSP in the sinusoidal compartment. The full liver toxicokinetic model developed in this laboratory (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) takes into consideration several processes not used explicitly in this work, including nonlinear transport at the sinusoidal and biliary membranes and metabolism. Each of these processes has the potential to saturate, producing non-linear effects on the net uptake rate. However, the experimental conditions in this work were chosen to minimize the likelihood of non-linear uptake. By keeping the BSP to albumin ratio low, the amount of free BSP in the sinusoids was kept low, which would tend to keep the membrane transport and subsequent processes operating at rates well below saturation.

The experiments and modeling presented here indicate that a low dissociation rate for a chemical-protein binding interaction can alter the kinetics of the clearance of a chemical by the perfused liver. When using a given set of perfused liver experiments or any limited set of experiments to extrapolate to a wide variety of doses and conditions, ignoring this dissociation limited effect can lead to inaccurate predictions.

Acknowledgments

The authors would like to thank Frank Dessauer and Marcia Feldmann for preparation of isolated perfused rat livers and Wayne Brashear, David Tsui and Eric Eldridge for HPLC analysis.

References

Baker, K. J., and Bradley, S. E. (1966). Binding of sulfobromophthalein (BSP) sodium by plasma albumin. Its role in hepatic BSP extraction. *J. Clin. Invest.* **45,** 281-297.

Fang, S. C. and Lindstrom, F. T. (1980). In vitro binding of ^{14}C-labeled acidic compounds to serum albumin and their tissue distribution in the rat. *J. Pharmacokin. And Biopharm.* **8,** 583-597.

Foy B.D., Toxopeus C., Frazier J.M. Kinetic Modeling of Slow Dissociation of Bromosulfophthalein from Albumin in Perfused Rat Liver: Toxicological Implications. *Toxic. Sci.* 1999 (in press).

Frazier J.M. Toxicol Methods 8, 257-285, 1998.

Gartner U, Goeser T, Wolkoff AW. Gastroenterology 113:1707-1713, 1997.

Geng WP, Schwab AJ, Goresky CA, and Pang KS. Hepatology 22: 1188-1207, 1995.

Goresky, C. A. (1964). Initial distribution and rate of uptake of sulfobromophthalein in the liver. *Am. J. Physiol.* **207,** 13-26.

Goresky, C. A. (1980). Uptake in the liver: The nature of the process. In *Liver and Biliary Tract Physiology*, (N. B. Javitt, Ed.). Vol. 21, pp. 65-101. University Park Press, Baltimore, MD.

Gumucio, D. L., Gumucio, J. J., Wilson, J. A. P., Cutter, C., Krauss, M., Caldwell, R., and Chen, E. (1984). Albumin influences sulfobromophthalein transport by hepatocytes of each acinar zone. *Am. J. Physiol.* **246,** G86-G95.

Koreniewski, C., and Callewaert, D. M. (1983). An enzyme-release assay for natural cytotoxicity. *J. Immun. Methods* **64,** 313-320.

Ockner, R. K., Weisiger, R. A., and Gollan, J. L. (1983). Hepatic uptake of albumin-bound substances: albumin receptor concept. *Am. J. Physiol.* **245,** G13-G18.

Orzes, N., Bellentani, S., Aldini, R., Simon, P., Ferretti, I., Lunazzi, G. C., Sottocasa, G. L., Tiribelli, C. (1985). Sex differences in the hepatic uptake of sulfobromophthalein in the rat. *Clin. Sci.* **69,** 587-593.

Ott, P., and Weisiger, R. A. (1997). Nontraditional effects of protein binding and hematocrit on uptake of indocyanine green by perfused rat liver. *Am. J. Physiol.* **273,** G227-G238.

Pfaff, E, Schwenk, M., Burr, R., Remmer, H. (1975). Molecular aspects of the interaction of bromosulfophthalein with high-affinity binding sites of bovine serum albumin. *Molec. Pharm.* **11,** 144-152.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* pp. 656-706. Cambridge University Press, Cambridge.

Schwenk M, Burr R, Schwarz L, Pfaff E. Eur J Biochem 64: 189-197, 1976.

Son, D. S., Osabe, M., Shimoda, M., Kokue, E. (1998). Contribution of alpha 1-acid glycoprotein to species difference in lincosamides-plasma protein binding kinetics. *J. Vet. Pharmacol. Ther.* **21,** 34-40.

Sorrentino, D., Zifroni, A., Van Ness, K., Berk, P. D. (1994). Unbound ligand drives hepatocyte taurocholate and BSP uptake at physiological albumin concentration. *Am. J. Physiol.* **266,** G425-G432.

Sultatos, L. G., Basker, K. M., Shao, M., Murphy, S. D. (1984). The interaction of the phosphorothioate insecticides chlorphyrifos and parathion and their oxygen analogues with bovine serum albumin. *Mol. Pharmacol.* **26,** 99-104.

van der Sluijs, P., Postema, B., and Meijer, D. K. F. (1987). Lactosylation of albumin reduces uptake rate of dibromosulfophthalein in perfused rat liver and dissociation rate from albumin *in vitro*. *Hepatology* **7,** 688-695.

Weisiger, R. A., Zacks, C. M., Smith, N. D., and Boyer, J. L. (1984). Effect of albumin binding on extraction of sulfobromophthalein by perfused elasmobranch liver: Evidence for dissociation-limited uptake. *Hepatology* **4,** 492-501.

Weisiger, R. A. (1985). Dissociation from albumin: a potentially rate-limiting step in the clearance of substances by the liver. *Proc. Natl. Acad. Sci., USA* **82,** 1563-1567.

Wyman, J., Stokes, J. S., Goehring, M., Buring, M., and Moore, T. (1995). Data collection interface for isolated perfused rat liver: recording oxygen consumption, perfusion pressure, and pH. *Toxicol. Meth.* **5,** 1-14.

Zhao, Y., Snel, C. A. W., Mulder, G. J., Pang, K. S. (1993). Localization of glutathione conjugation activities toward bromosulfophthalein in perfused rat liver. *Drug Metab. Dispos.* **21,** 1070-1078.

GOOD NEWS: WORK SAMPLES ARE (ABOUT) AS VALID AS WE'VE SUSPECTED

Charles E. Lance
Professor
Department of Psychology
The University of Georgia
Athens, GA 30602-3013

Final Report for Summer Research Extension Program Subcontract 980842.

This research was supported by the Air Force Office of Scientific Research, Bolling AFB, Washington, DC, and The University of Georgia, Athens, GA.

February, 2000

GOOD NEWS: WORK SAMPLES ARE (ABOUT) AS VALID AS WE'VE SUSPECTED

Charles E. Lance
Professor
Department of Psychology
The University of Georgia

Abstract

Data obtained on over 1,500 first-term U.S. Air Force enlisted personnel indicated that work sample administrators' global ratings of work sample performance substantially reflect actual ratee behavior in the work sample, and not potentially biasing factors (e.g., race, gender, amount of recent experience), supporting the "folk wisdom" that these global performance judgments are, in fact, valid and unbiased measures of performance. Good news!

GOOD NEWS: WORK SAMPLES ARE (ABOUT) AS VALID AS WE'VE SUSPECTED

Charles E. Lance

A work sample may be defined as "...a measure of performance on a structured task that is directly reflective of the type of behaviors required in the job situation" (F.D. Smith, 1991, p. 28). As such, work sample measures may be distinguished from other related performance measurement approaches such as (a) trainability tests, which include a specified time to learn the task to be performed (Robertson & Downs, 1989), (b) situational judgment tests, which typically require the examinee to respond to a hypothetical situation, and which may be administered either in an oral or written mode, and (c) job knowledge tests, which assess declarative (as opposed to procedural) aspects of performance, and which usually are administered in written form. For many years, the work sample has been touted as an effective approach to the measurement of work-related behaviors for the purposes of predicting subsequent on-the-job performance, assessing training effectiveness, and measuring current job proficiency (F. D. Smith, 1991). Results of Terpstra's (1996) recent survey document the long-held and widespread <u>belief</u> in the effectiveness of work samples among human resource executives.

The effectiveness of the work sample as a <u>predictor</u> of job success has been documented in several reviews. For example, Asher and Sciarrino (1974) classified work samples as either motor ("...if the task was a physical manipulation of things..." p. 519) or verbal ("...if there was a problem situation that was primarily language-oriented or people oriented." p. 519), and found motor work samples to be second only to biodata in terms of their predictive efficiency; verbal work samples were somewhat less predictive of performance. Later reviews confirm these basic findings: work samples are among the most valid predictors of job performance, having mean validities in the .40s to .50s (Hunter & Hunter, 1984; Robertson & Kandola, 1982; Schmidt & Hunter, 1998; Schmitt, Gooding, Noe, & Kirsch, 1984). M. Smith's (1994) theory of the validity of predictors of job performance suggests reasons <u>why</u> work samples are valid. First, work samples are very effective at assessing specific abilities and specialized skills that are required for the performance of particular jobs. Second, work samples are thought of as objective measures of samples of behavior that are highly representative of actual job duties. Thus, from M. Smith's

theory, the high validity of work samples is seen to arise from the objective assessment of representative job duties that are required for successful job performance in specific jobs.

Work samples have also been held in high regard as criterion measures (Borman, White, & Dorsey, 1995; Kavanagh, Borman, Hedge, & Gould, 1987). For example, Borman and Hallum (1991) noted that "...some researchers have maintained that work samples...are the highest fidelity performance-measurement method available and that they provide the most valid indication of 'actual' performance" (p. 11). Some have even suggested that alternative measures of performance (e.g., performance ratings) might be validated in terms of their relationships with work sample measures (Wigdor & Green, 1991a).

Granted, work sample performance measures are high fidelity measures of "can-do" aspects of job performance (versus "will-do" aspects, see Borman et al., 1995; Borman & Motowidlo, 1993, 1997; Borman, White, Pulakos, & Oppler, 1991; DuBois, Sackett, Zedeck, & Fogli, 1993; Motowidlo, Borman & Schmit, 1997; Motowidlo & VanScotter, 1994; Sackett, Zedeck, & Fogli, 1988). However, they are not without their limitations. First, they can be time consuming, labor intensive, and expensive to develop and operate (Asher & Sciarrino, 1974; Hedge & Teachout, 1992; Hunter & Hunter, 1984; F. D. Smith, 1991). Second, although tasks included within work sample test batteries usually represent corresponding on-the-job elements with high fidelity, a relatively small range of job tasks is usually included in them due to the time and expense in developing and operating them. Thus except for some highly specialized jobs (e.g., life guard, toll taker, raisin washer), work samples may suffer more from criterion deficiency (Thorndike, 1949) as compared to alternative criterion measures. Third, work sample measures tap into only a subset of the complete criterion construct domain. For example, work samples reflect the job specific task proficiency factor in Campbell's (1990; see also Campbell, McHenry, & Wise, 1990) criterion domain taxonomy, but not other factors such as facilitating peer and team performance, supervision, and written and oral communication (except as related to specific job requirements). Finally, there is some question as to whether the criterion measures typically obtained from work samples (administrators' global ratings of work sample task performance) are really as objective as has been presumed. This is the subject of the present study.

F. D. Smith (1991) characterized three typical approaches to scoring performance in a work sample. In one, a global rating approach, the work sample administrator observes examinee

performance in the work sample and rates the examinee's performance on global, usually Likert-type, scales with anchors such as "Unsatisfactory" to "Exceeds performance standards." This approach is commonly used in scoring work sample task performance (F. D. Smith, 1991), and is also the approach that is most often used in the assignment of assessment center post-exercise dimension ratings (Klimoski & Brickner, 1987; Thornton, 1992). Ratings on a number of such scales (e.g., representing different performance dimensions) may be averaged to form an overall score for each work sample task, and ratings may be averaged across tasks to form an overall work sample test battery score. In a second approach, the work sample administrator is provided with behavioral recording forms which list specific examples of good and poor performance (developed by subject matter experts - SMEs) in the work sample that are intended to guide the administrator in observing examinee behaviors and in making summary, global ratings of examinee task performance. Thus in this approach, the work sample administrator still makes only a single global rating, but with the assistance of behavioral exemplars to guide the global performance judgment. Finally, work samples may be scored using a behavioral checklist. In this, probably the least often used approach, the work sample administrator indicates which of a number of prespecified task steps were completed either correctly or incorrectly by the examinee (see e.g., Brugnoli, Campion, & Basen, 1979; Campion, 1972). In this approach, work sample performance is usually scored as some form of percentage of task steps completed correctly for each task. Once again, overall work sample performance may be scored by computing an aggregate score across all tasks included in the work sample test.

Nearly 25 years ago, P. C. Smith (1976) distinguished between "hard" (i.e., objective) and "soft" (i.e., subjective) criterion measures in terms of the extent to which they involve subjective judgment. Noting that work sample administrator judgment is required in each of the work sample scoring strategies described by F. D. Smith (1991), it could be argued that none of these scoring schemes is entirely "objective." Despite the high fidelity of work sample tasks, work sample administrators still must make "clinical" global performance judgments of task performance based on their observations of examinee behaviors in the first scoring scheme (i.e., the global rating approach, though these may be aided with behavioral recording forms, as in the second scoring option, the behavior recording forms approach). Even the use of behavioral checklists (as in the third scoring scheme) may require judgment as to whether individual task steps are completed

correctly or incorrectly. For example, Borman and Hallam (1991) found there may not only be disagreement among work sample administrators as to whether particular executions of task performance steps are correct or incorrect, but there may even be more fundamental disagreement among SMEs as to what target behaviors should be scored as correct or incorrect.

Aside from Borman and Hallam's (1991) study, there has been almost no research on the validity of work sample administrators' performance judgments. One exception was Hedge and Teachout's (1992) study of the convergence between work sample tasks as administered in a hands-on mode as compared to an interview mode. Work sample validity has been inferred from the fidelity with which actual job tasks are represented in the content of work sample test batteries. But replicating job functions in a high-fidelity simulated test environment does not insure that the examinee will be evaluated objectively in this environment. Human (i.e., work sample administrator) judgment, has been shown to be subject to a number of biases (Borman, 1991). In fact, performance ratings have a history of having been presumed to be contaminated with various judgment and rating errors (Landy & Farr, 1980; Saal, Downey, & Lahey, 1980) and much of the historically relevant models of performance rating processes seem preoccupied with identifying and describing all the ways in which these judgment and rating processes are prone to error (e.g., Ilgen & Feldman, 1983; Landy & Farr, 1980; Wherry & Bartlett, 1982). Consequently, it may well be expected that errors and biases in work sample administrator judgments (qua performance ratings) may be significant factors in the measurement of work sample performance, and particularly if only global task performance ratings are made.

The purpose of this study was to provide an empirical assessment of whether work sample administrator global ratings of examinee performance on work sample tasks substantially reflect actual examinee performance in the work sample (as is generally presumed), or whether they may be biased by factors that are not directly related to examinee performance in the work sample. The particular work samples reported here afforded a unique opportunity to test these ideas, as data were collected on (a) work sample administrators' global task performance ratings (i.e., as in the global rating approach described by F. D. Smith, 1991), (b) whether discrete task steps that comprised the work sample tasks were completed correctly or incorrectly (i.e., as in the behavioral checklist approach described by F. D. Smith, 1991), and (c) a number of additional factors related to task performance (e.g., time to complete work sample tasks, previous experience performing the

tasks, examinee demographic characteristics, etc. These are described in greater detail later.). We predicted that if work sample administrators' global task performance ratings are as valid as they have been presumed, then they should substantially reflect actual examinee behavior in the work sample and not the influences of other factors that are less directly related to task performance. To our knowledge, only one previously published study has attempted to address this issue (Brugnoli et al., 1979), in which it was found that global ratings, but not behavioral checklist measures of work sample task performance was subject to racial bias. However, this was a small-sample (N = 46) laboratory study in which work sample performance was depicted only in brief videotaped segments that showed only the examinees' arms and hands. So, in addition to the main focus of our study, we also were interested in the extent to which Brugnoli et al's. (1979) results would be replicated in a much larger sample and more ecologically valid measurement context.

Summary and Specific Predictions

Work samples have long been touted as objective performance measures, yet very little research has investigated their ostensible objectivity. That is, even in high-fidelity measurement situations, work sample administrator global task ratings, like supervisory performance ratings, may reflect non-performance-based information (Lance, Woehr, & Fisicaro, 1991) as well as performance-based information available to the work sample administrator in the test situation. We predicted that if global work sample ratings are as valid as they have been presumed, then they should substantially (and perhaps exclusively) reflect actual examinee behavior in the work sample (i.e., percentage of task steps completed correctly) and not other factors that are peripherally related to work sample performance. Thus the present research was designed as a policy-capturing study (Cooksey, 1996; Hoffman, 1960) of the extent to which information which is available to the rater during work sample administration (information which relates both to performance-related and performance-irrelevant factors), combines to affect administrators' overall judgments of work sample task performance. To date, there has been almost no research on this question.

Method

Study Context

Data were collected as part of a large-scale Joint-Service Job Performance Measurement (JPM)/Enlistment Standards project conducted by the U.S. military in the late 1980s and early 1990s (Wigdor & Green, 1991a, 1991b). The major purposes of this project were to link

enlistment standards to on-the-job performance and to explore alternative technologies for measuring job performance. The conceptualization, design, and execution of the JPM Project has been discussed extensively elsewhere (Hedge & Teachout, 1986, 1992; Kavanagh et al., 1987; Lance, Teachout, & Donnelly, 1992; Laue, Hedge, Wall, Pederson, & Bentley, 1992; Ree, Earles, & Teachout, 1994; Teachout & Pellum, 1991). Thus only the particular aspects of the JPM Project that are relevant to the present study are highlighted here.

Samples

Samples were obtained from eight U.S. Air Force (USAF) specialties (AFSs) selected for inclusion in the JPM Project. These included Aircrew Life Support Specialist, $n = 229$; Air Traffic Control Operator, $n = 190$; Precision Measurement Equipment Laboratory Specialist, $n = 140$; Avionic Communication Specialist, $n = 98$; Aerospace Ground Equipment (AGE) Mechanic, $n = 269$; Jet Engine Mechanic, $n = 255$; Information Systems Radio Operator, $n = 155$; and Personnel Specialist, $n = 200$. These AFSs were selected to be representatative of (a) the relatively more populous jobs in the enlisted occupational classification structure in existence at the time of data collection, (b) varying levels of occupational learning difficulty (see Burtch, Lipscomb, & Wissman, 1982; Mumford, Weeks, Harding, & Fleishman, 1987; Weeks, 1984), and (c) existing accession and classification policies based on mechanical, administrative, general, and electronic (MAGE) aptitude requirements (Department of Defense, 1984; i.e., two AFSs were chosen to represent each of the four MAGE aptitude areas).

Work Sample Task Selection

A number of criteria were used to select tasks for inclusion in each JPM AFS's work sample test battery. First, occupational survey (i.e., job analysis) data were analyzed to identify those tasks that were most widely performed by first-term incumbents in the respective AFSs. Second, tasks were selected from among the most frequently performed tasks to insure that most examinees would have some experience performing tasks included in the work sample test battery previously on the job. Third, tasks were also selected so as to reflect a range of task learning difficulties. Specifically, 40% of the work sample test battery tasks were sampled from the fourth quartile of task learning difficulty (i.e., the most difficult), 30% from the third quartile, 20% from the second quartile, and 10% from the first (least difficult) quartile, in order to reduce the likelihood of ceiling effects in work sample performance. Fourth, candidate tasks were reviewed by SMEs in "task

validation workshops" (see Laue et al., 1992) to insure that constituent task steps were observable and that they could be scored unambiguously as being completed correctly or incorrectly. The purposes of this selection criterion were to insure that (a) discrete task steps were directly identifiable and observable by work sample administrators, and (b) performance on each task step could be scored as correct or incorrect according to specified criteria, thus minimizing potential ambiguities in scoring work sample test items that Borman and Hallam (1991) had identified earlier. Table 1 shows one example of a work sample task included here (<u>Installation of engine pressure ratio probes</u> for the Jet Engine Mechanic AFS) and constituent task steps, each of which was scored on a correct/incorrect or "go/no-go" basis. Candidate tasks whose steps were either not easily observable or scorable on to a "go/no-go" basis were replaced with alternate tasks. Combined with the extensive work sample administrator training that was given prior to administration of the work sample tasks (described below), these task selection criteria insured that evaluation of work sample performance at the step-level was as objective as may be possible in an operational work sample test battery.

Altogether, between 20 and 46 tasks per AFS were selected for the work sample task batteries. In most AFSs, some tasks were widely performed by all incumbents (referred to as "Phase I" tasks), while others ("Phase II" tasks) were performed only by incumbents in particular functional areas. For example, Jet Engine Mechanic Phase I tasks were commonly performed by all Jet Engine Mechanics, but Phase II tasks varied as a function of the particular type of jet engine that the incumbent serviced. In order to maximize sample sizes and to insure that the work domains represented here were content valid for all members of respective AFSs (and not merely a more specialized subset), only Phase I tasks were included in this study, resulting in the retention of between 8 and 31 tasks per AFS for analysis (see Table 3, below).

<u>Work Sample Administrator Training</u>

Work sample tests were administered by active-duty or recently retired noncommissioned officers from the respective AFSs. Administrators received 1 - 2 weeks of intensive training in the observation and scoring of the work sample tests. Training included procedures for work sample administration, observation of examinee performance, and work sample scoring procedures. Training methods included lecture and discussion, role playing, and viewing and discussing videotaped target task performances. Videotaped task performances were scripted to reflect both

correct and incorrect step-level performances, and to establish a common frame of evaluative reference among the test administrators. After they viewed and scored the videotaped task performances, trainers and work sample administrator trainees discussed in detail the key behaviors depicted in the videotaped performances to reach concensus on what behaviors would be scored as correct and incorrect subsequently during test administration. Inter-administrator reliability was estimated at .81 to .98 (see Hedge, Dickinson, & Bierstedt, 1988 and Hedge & Teachout, 1992 for additional details).

Procedure

Upon arriving at the test station, examinees were briefed as to the general purpose of the work sample test and were administered an appropriate work sample test battery. Examinees were instructed and encouraged to do their best on each work sample task. Testing required 4 - 8 hours per examinee. For each work sample task, the work sample administrator recorded (a) incumbent-estimated number of times s/he had performed the task previously on the job ("Number of Times Performed"), (b) how long it had been (in weeks) since s/he had last performed the task ("Last Time Performed"), and (c) time of day at the beginning of the task administration. Next, the administrator administered the work sample task to the examinee, observed examinee task performance, and recorded whether each task step was completed correctly.[1] Third, the administrator recorded the time at the completion of the task and the total time required to complete the task ("Time Required"). Finally, the administrator completed a global rating of task performance ("Overall Performance:" "1 = Far below the acceptable level of proficiency," to "5 = Far exceeded the acceptable level of profiency").[2] These four steps were repeated for each task in the work sample test battery. However, the second step (i.e., task administration) occurred in two different modes: hands-on and interview. In the hands-on mode, examinees were instructed to perform the task as they would on the job, and were allowed access to technical manuals and other written materials as they would ordinarily on the job. In the interview mode, examinees were asked to describe the steps necessary for task completion in a "show and tell" manner, but without the aid of technical manuals or other information (see Hedge & Teachout, 1992). Some work sample tasks were administered in the hands-on mode only, some in the interview mode only, and some in both (referred to by Hedge & Teachout, 1992 as "overlap tasks"). For overlap tasks, the interview

mode of administration always preceded the hands-on administration of the work sample task. We included both hands-on and interview work sample tasks for analysis.

Measures

Overall Performance (OAP) was the work-sample administrator's global 5-point rating of work sample task performance, and was the primary criterion variable in this study. Note that this measure is typical of many work sample task-level performance measures, and exemplifies the overall work sample task ratings obtained in the global rating and behavioral recording forms approaches to scoring work samples described by F. D. Smith (1991).

Percent Steps Correct (%Correct) was measured as an unweighted percentage of task steps completed correctly as recorded by the work sample administrator. Note that this measure is typical of the behavioral checklist approach to scoring work sample task performance as described by F. D. Smith (1991). As such, it provides perhaps the closest possible link, particularly with the task selection and administrator training safeguards implemented in the work sample test batteries reported here, between measured task performance and actual examinee behavior in the work sample situation. The high interscorer (i.e., shadow score) reliabilities reported earlier also are testimony to the objectivity of these measures. We predicted that %Correct would be positively related to OAP, and if OAP-type ratings are as objective and valid as has been presumed, that %Correct would account for substantially all of the predictable variance in OAP. Otherwise, we expected that OAPs might also reflect substantial influences of one or more of the following variables which relate more peripherally to actual performance in the work sample.

Number of Task Steps (#STEPS). As mentioned earlier, each work sample task consisted of a number of discrete task steps which were identified from the respective AFSs' technical and training manuals ("technical orders"). The number of constituent task steps ranged between 2 and 47. #STEPS can be considered as an indicator of task complexity. We expected that significant OAP -- #STEPS relationships would be negative, that is, that performance would generally be rated lower on more (versus less) complex tasks, as more complex tasks would be generally perceived as being more difficult.

Time to Complete Task (TIME), measured in minutes, was the difference between the work sample task finish time and start time. For cases in which the examinee did not finish the task within the pre-established time limit, TIME was set equal to the time limit. We expected that, all

other things equal, OAPs would be higher for quicker (and perhaps more expertly executed, versus slower) task performances.

Last Time Performed (LTP) was computed as the number of weeks since the task had last been performed as part of the examinee's regular job duties. Thus LTP indicated the length of the interval in between the time the task was last performed and the time it was tested in the work sample (Lance, Parisi, Bennett, Teachout, Harville, & Welles, 1998). All other things equal, we expected higher OAPs for cases in which the task had been performed on the job more recently, as more recent experience might be expected to facilitate task performance.

Number of Times Performed (NTP) were incumbents' reports of the number of times they had previously performed the task on the job as part of their regular job duties. Previous research (e.g., Lance, Hedge, & Alley, 1989; Lance et al., 1998) has found that NTP is markedly positively skewed and multimodal. Thus we transformed it (as in previous studies) as 1 = Never performed, 2 = 1 to 10 times performed previously, 3 = 11 to 20 previous performances, 4 = 21 to 50, 5 = 51 to 100, 6 = 101 to 800, and 7 = 801 to 999 previous performances ("999" indicated that the examinee had performed the task so often that they could not estimate the number of previous performances). We expected positive OAP -- NTP relationships, that is, higher OAPs for cases in which the task had been performed often previously, as more experienced examinees might be expected to perform more effectively than less experienced ones.

Examinee Motivation (MOT) to perform effectively in the work sample test was measured as a composite of six items anchored by 5-point Likert-type scales. These items were included on a questionnaire that was completed by the work sample examinee immediately after completing the work sample test battery. Example items included "Did you feel that it was important to perform well on the (work sample) test?" and "How motivated were you to perform to the best of your ability on the (work sample) test?" Standardized coefficients alpha ranged between .81 and .86 across AFSs[3]. We predicted positive OAP -- MOT relationships on the basis that more motivated performance may serve as a cue to performance effectiveness (Martell, Guzzo, & Willis, 1995).

Demogaphic Variables. Sex was scored as Male = 1 and Female = 0. Personnel records included three racial codes for "White," "Black," and "Other." We recoded race as two binary variables: White (= 1, 0 = Nonwhite), and Black (= 1, 0 = Nonblack). We included these factors because gender and racial biases in performance measures have been found previously (e.g.,

Brugnoli et al., 1979; Ford, Kraiger, & Schechtman, 1986; Hamner, Kim, Baird, & Bigoness, 1974; Tosi & Einbender, 1985), although their effects are often minimal or nonexistent under performance measurement conditions such as in the present study (Pulakos, White, Oppler, & Borman, 1989; Tosi & Einbender, 1985; Sackett & DuBois, 1991).

Data Analyses

We performed two complementary sets of analyses. Both were aimed at determining what information that is available during work sample administration impacts administrators' OAP ratings. That is, both analytic approaches were directed toward capturing work sample administrators' OAP rating policies (Cooksey, 1996). In the first, we used ordinary least squares (OLS) multiple regression to regress the global task performance rating (OAP) for each task on %Correct in the first step, and in the second step, also on TIME, LTP, NTP, MOT, Sex, White, and Black. We entered TIME, LTP, NTP, MOT, Sex, White, and Black after entering %Correct, because some of these variables could be considered as performance determinants (e.g., task experience [indexed by NTP], and examinee motivation [MOT] should, theoretically, enhance task performance). Thus the effects of these variables on OAP should be considered as peripheral only to the extent that their effects on actual work sample task performance have already been controlled. Thus we controlled for these effects by entering %Correct into the policy-capturing equation first, followed by the remaining variables in step 2. We evaluated the change in $\underline{R}^2$ (i.e., $\Delta\underline{R}^2$) from the first to the second step to investigate the statistical and practical significance of the variables included in the second step.

Altogether, we performed 134 such hierarchical regressions corresponding to the total number of Phase 1 tasks included in all eight AFSs. Sample sizes for each regression equation varied across AFSs (as were reported earlier). Support for the validity of the OAPs would be obtained if %Correct accounted for a substantial proportion of variance in OAP, and if the remaining variables accounted for very little variance in OAP beyond that which was accounted for by %Correct. Bias in OAPs would be indicated to the extent that one or more of the additional variables accounted for a substantial proportion of variance in OAP beyond that accounted for by %Correct.

The second analytic strategy combined data for all 134 tasks into a single stacked multi-level data set. This data set was multi-level in the sense that variables were operationalized at

varying levels of specificity. For example, the study's dependent variable (OAP) indexed the $\underline{i}$th examinee's ($\underline{i} \rightarrow \underline{N}_k$ as reported earlier for each of the $\underline{k} \rightarrow \underline{K} = 8$ samples) performance on the $\underline{j}$th work sample task ($\underline{j} \rightarrow \underline{J}_k$, $\underline{J}_k$ ranged between 8 and 31). Thus, the effective sample size was $\Sigma(\underline{N}_k{*}\underline{J}_k) = 14{,}965$ after the deletion of missing data. %Correct also varied both across examinees and tasks, as did TIME, LTP, and NTP. Thus, OAP, %Correct, TIME, LTP, and NTP were task x examinee-level variables. On the other hand, #STEPS varied across the $\underline{j}$ tasks, but was constant for all $\underline{N}_k$ performers of the $\underline{j}$th task. Thus #STEPS was a task-level variable. Finally, examinee motivation (MOT), Black, White, and Sex were three examinee-level variables, as they varied appropriately across the $\underline{N}_k$ examinees, but were constant for the $\underline{i}$th examinee across his/her performance of the $\underline{J}_k$ tasks attempted in the work sample test battery.

We also explored possible interactions between %Correct and an additional binary variable indicating whether the task was administered in the interview (=0) or hands-on (=1) mode ("H/I"), and the additional predictors, as Hedge and Teachout (1992) indicated that mode of administration may impact factors related to task performance. To do this we first centered %Correct, H/I and the remaining predictors (i.e., to a mean of zero), and then formed cross-products between %Correct and H/I and the additional predictors (e.g., %Correct x #STEPS, H/I x LTP, etc.). Finally, we entered these cross-product terms into the OAP regression equation in a third step. However, since we had no a priori predictions regarding interaction effects, we entered the cross-product terms using forward selection with an $\alpha < .05$ entry criterion. Finally, results reported later suggested that the form of the %CorrectxTIME interaction might vary between hands-on and interview tasks. We tested this by entering the 3-way H/Ix%CorrectxTIME interaction in a fourth step in the regression model.

Results

Table 2 shows study variables' descriptive statistics and intercorrelations for all AFSs combined.[4] Mean OAP and %Correct values indicated the absence of ceiling effects and their SDs indicated that range restriction was not a problem. The mean NTP indicated that, on the average, examinees were experienced performing the tasks on which they were examined in the work sample test battery, but the mean LTP indicated that, on the average, it had been about 3 1/2 months since they had last performed the tasks included in the work sample on the job. MOT scores generally

indicated that examinees were in fact motivated to perform well in the work sample. Finally, data in Table 2 show that the total sample was 80% White, 13% Black, and 85% Male.

Table 2 also shows that, as predicted, OAP was positively correlated with %Correct. However, NTP, TIME, LTP, MOT, and #STEPS also were significantly correlated with OAP, and in the hypothesized directions. Notably, correlations among most predictor variables were quite low (but statistically significant, due to the extremely high power afforded by the combined samples' size), and exceptions are easily understood. For example, (a) r(TIME,#STEPS) = .52 indicates that, on the average, it takes longer to perform tasks that have more constituent task steps, (b) r(NTP,TIME) = -.18 indicated some tendency for more experienced examinees to perform task more quickly, (c) r(NTP,LTP) = -.37 indicated that more experienced examinees also tended to have more recent experience on tasks in the work sample test battery, and (d) r(TIME,H/I) = .38 indicated that it took examinees somewhat longer (on the average) to actually perform hands-on tasks than it did for them to explain how they would perform tasks as administered in the interview mode. Also notable is the fact that correlations between demographic and more substantive variables are near zero, and many are statistically nonsignificant. This reinforces previous research indicating that when racial and gender biases are found, their effects are often quite small (Pulakos et al., 1989; Tosi & Einbender, 1985; Sackett & DuBois, 1991). Finally, correlations with H/I indicated that there was some tendency for examinees to obtain higher performance scores on hands-on tasks as compared to tasks administered in the interview mode. Tables 3 through 5 address the study's main questions more directly.

Table 3 shows the percentages of regression equations from the first set of analyses in which each variable was a statistically significant (i.e., $p < .05$) predictor of OAP. Numbers outside (inside) parentheses indicate the percentages of times that each predictor was statistically significant and was signed in the predicted (opposite) direction. For example, the first row of Table 3 shows that of the 19 regression equations for the Avionics Communication sample (i.e., one equation for each work sample task), %Correct was a statistically significant (and properly signed) predictor of OAP in 100% (i.e., all 19) of the equations; NTP was a statistically significant (and properly signed) predictor in 15.8% of the equations; TIME was a statistically significant (and properly signed) predictor in 36.8% of the equations but a statistically significant (and oppositely signed) predictor in 10.5% of the equations, and so forth. The last row summarizes the mean percentages

across all samples. The first column of Table 3 shows that %Correct was a significant predictor of OAP in nearly every regression equation, and in no case was the effect of %Correct on OAP estimated to be statistically significant and negative. The last entry in the second column indicates that NTP was a statistically significant (and properly signed) predictor of OAP in 9% of the estimated equations, but in 1.5% of the equations the coefficient was statistically significant but negative (contrary to predictions). Table 3 also shows that overall, LTP, MOT, Black, White, and Sex were "significant" predictors of OAP at, or well, below chance levels. Interestingly however, TIME was a significant predictor of OAP in a total of 29.1% of the regression equations, but in many cases (17.9% of the equations) its coefficient was negative (as was predicted) and in others (11.2% of the equations), the coefficient's sign was positive.

To try to pinpoint the reason for why TIME's coefficient was sometimes negative and sometimes positive, we summarized regression equations separately for hands-on and interview work sample tasks. For the 25 equations in which TIME (in addition to %Correct) was a statistically significant predictor of hands-on task OAPs, its coefficient was negative (as was predicted) in 20 (80%) of them. However, for the 14 equations in which TIME was a statistically significant predictor of interview task OAPs, its coefficient was positive (opposite to that predicted) in 10 (71%) of them. This difference in patterns of relationships between OAPs and TIME between hands-on and interview tasks was itself statistically significant: $\chi^2(1) = 10.06$, $p < .01$. That is, controlling for %Correct, administrators gave somewhat higher OAP ratings for quicker performances in hands-on tasks, and higher OAP ratings for slower performance in interview tasks. We interpret this as indicating that administrators gave "extra credit" for quickly and smoothly-executed hands-on performances, and for more detailed and thorough (though slower) "show-and-tell" explanations of task performance in interview tasks.

On the whole, however, %Correct overshadowed every other predictor in accounting for variance in OAP ratings. This conclusion is further reinforced in Table 4 which shows mean R^2 and β (i.e., standardized regression coefficient) values ($\pm$ 1 SD) calculated across the 134 regression equations (values were converted to zs, averaged, and backtransformed to R^2s and βs). The mean R^2 (.54) approaches the reliability of global performance ratings as cited by Viswesvaran, Ones, and Schmidt (1996). That is, %Correct accounts for nearly all of the variance in OAP that could potentially be accounted for, given Viswesvaran et al.'s (1996) estimates of the reliability of

performance ratings. Second, %Correct accounts for 88% of the variance in OAP that, on the average, is accounted for in the full regression equations (i.e., $\beta^2/\underline{R}^2 = .69^2/.54 = .88$). Thus OAPs substantially reflect the influence of examinee behavior in the work sample (%Correct) and not the effects of additional factors that are more peripherally related to performance in the work sample.

Results from the second set of analyses complement and extend these findings. The overall β for %Correct shown in Table 5 ($\beta = .759$) is on the same order as the mean β for %Correct reported in Table 4 (.69). And although the variables added in Step 2 of the regression model explained a statistically significant proportion of variance in OAP above and beyond that which was predicted by %Correct ($\Delta\underline{R}^2 = .006$, $\underline{F} = 24.92$, $\underline{p} < .001$), %Correct alone accounted for 99% of the variance explained on the basis of the Step 2 regression equation (i.e., .577/.583 = .9897). Nevertheless, effects of the additional variables, although small, were in the predicted directions. All other things equal, OAPs were somewhat higher for (a) examinees who reported as having been more motivated to perform well in the work sample (effect of MOT), (b) tasks with fewer steps (#STEPS, i.e., simpler tasks), (c) examinees who had performed the task on the job more recently (LTP), (d) examinees who had performed the task more often (NTP), and (e) examinees who performed tasks more quickly (TIME). There also were small effects favoring Blacks and Whites (versus "Other" groups) and against Males. However, all of these additional effects (i.e., beyond the effect of %Correct on OAP) must be interpreted in the contexts that (a) collectively, they account for only about 1% of the variance explained on the basis of the Step 2 regression model, and (b) these effects would likely remain undetected except for the extremely high statistical power afforded here by the large effective sample size ($\underline{N} = 14{,}965$).

A number of statistically significant 2-way interaction effects also were detected which, collectively, accounted for an additional 1.1% ($\underline{F} = 57.18$, $\underline{p} < .001$) of the variance in OAP. Again, most of these effects were small, and were detectable only by virtue of the extremely high power afforded by the large effective sample size in this second set of analyses. The %Correct x TIME interaction indicated that OAPs were low for low values of %Correct regardless of the amount of time taken to perform the task, but for higher values of %Correct, OAPs were higher for task performances that were executed more quickly than for slower task executions – administrators "gave extra credit" to effective task performances that were also executed quickly. The remaining interactions with %Correct followed the same general pattern: administrators "gave extra credit"

for effective task performances (i.e., high values of %Correct) (a) that occurred on more complex (more task steps) versus simpler tasks (fewer task steps, the PCx#STEPS interaction), (b) for individuals who had performed the task more often (vs. less often) previously (the PCxNTP interaction), and (c) for individuals who had performed the task relatively recently (i.e., the PCxLTP interaction). Lastly, the 2-way H/IxTIME interaction indicated a positive relationship between TIME and OAPs (longer performance times were associated with higher ratings) for tasks administered in the interview mode, while this relationship was nil for tasks administered in the hands-on mode.[5]

Finally, Table 5 shows that a 3-way interaction was supported between H/I, %Correct, and TIME which accounted for an additional .2% of the variance in OAPs. Consistent with findings from the first set of analyses, this 3-way interaction indicated that (a) for tasks administered in the interview mode, administrators gave somewhat higher ratings for effective task performances (high %Correct) when time to perform the task was longer (versus shorter), but (b) for tasks administered in the hands-on mode, administrators gave somewhat higher ratings for effective task performances when time to perform the task was shorter (versus longer). These results complement earlier findings indicating that administrators gave "extra credit" for more detailed and thorough (though slower) "show-and-tell" explanations of task performance in interview tasks, and for quickly and smoothly-executed performances in hands-on tasks.

Supplementary Analyses

So far, results support the idea that overall work sample task performance ratings (OAPs) substantially reflect actual examinee behavior in the work sample (%Correct) and not other, more peripheral factors (although these factors were shown to have predictable, albeit subtle, effects on OAPs). However, it could be argued that the strong and consistent %Correct – OAP relationships reported in Tables 2 through 5 reflect nothing more than consistent rater biases such as general impression halo error (Lance, LaPointe, & Stewart, 1994). That is, the observed %Correct – OAP relationship could be inflated simply because the work sample administrator applied the same (set of) bias(es) in making correct/incorrect step-level performance judgments (reflected in %Correct) and subsequent overall task performance judgments (OAPs). We tested this possibility using the shadow-scored data referred to earlier.

As we mentioned earlier, work sample performances for a relatively small number of examinees were scored concurrently and independently by a second work sample administrator (the shadow scorer) in addition to the one who actually administered the work sample test to the examinee. The number of examinees for which shadow score data were obtained were, for Aircrew Life Support Specialist, $\underline{n} = 8$; Air Traffic Control Operator, $\underline{n} = 18$; Precision Measurement Equipment Laboratory Specialist, $\underline{n} = 29$; Avionic Communication Specialist, $\underline{n} = 20$; Aerospace Ground Equipment (AGE) Mechanic, $\underline{n} = 14$; Information Systems Radio Operator, $\underline{n} = 20$; and Personnel Specialist, $\underline{n} = 17$.[6]

We re-ran analyses reported in Table 5 using shadow-scored %Correct (%Correct-Shadow) in lieu of the administrator's own %Correct scores as the primary indicator of actual examinee behavior in the work sample. Note that if findings reported in Table 5 substantially reflected same-source bias effects on %Correct and OAP scores, using %Correct-Shadow in lieu of %Correct would substantially reduce relationships found earlier, as %Correct-Shadow scores were obtained independently of OAP ratings. On the other hand, if using %Correct-Shadow scores substantially replicated earlier findings, this would underscore the veridicality of %Correct scores and earlier results indicating that OAPs substantially reflect actual examinee behavior in the work sample, and not potentially biasing factors.

Regression re-analysis results using %Correct-Shadow are reported in Table 6. Unlike results in Table 5, many of the predictor variables' effects were no longer statistically significant, owing to the substantial reduction in sample size and corresponding loss in statistical power. Nevertheless, the negative relationship between TIME and OAPs, and the PCSxTIME and PCSx#STEPS interaction effects on OAPs found earlier were replicated. However, the key findings were that, as in Table 5, (a) %Correct-Shadow, scored independently of OAPs, had a substantial and statistically signifcant impact on OAPs, and (b) %Correct-Shadow alone accounted for 97% of the variance that was accounted for in the full regression model (i.e., $\beta^2/\underline{R}^2 = .675^2/.471 = .966$). These results replicate earlier findings indicating that OAPs substantially reflect examinee behaviors actually exhibited in the work sample (as scored independently from OAPs), and not the effects of other factors that are more peripherally related to performance in the work sample. And, in hindsight, these results are not surprising since the correlation between %Correct and %Correct-Shadow ($\underline{r} = .954$, $\underline{N} = 2268$, $\underline{p} < .001$) indicated near perfect interscorer agreement. Together,

these supplementary results discredit the possible interpretation that %Correct – OAP relationships merely reflect consistency in work sample administrator bias across step-level and overall work sample task performance judgments.

Discussion

Combined, results in Tables 2 through 6 indicated that work sample administrator OAP ratings (a) substantially reflect the influence of examinee behaviors exhibited in the work sample, as indexed by %Correct (and %Correct-Shadow), (b) do not reflect racial or gender biases of any practical consequence, (c) are largely independent of potentially biasing effects of administrator prior knowledge of previous experience (indexed by NTP), recent experience (indexed by LTP) and possible performance-cue effects of ratee motivation (MOT), but (d) may reflect subtle stylistic aspects of performance (automaticity of task execution or thoroughness of explanation) that are not captured in a simpler count of the number of task steps that were completed correctly (differential effects of TIME on OAPs for hands-on vs. interview tasks). Thus in one sense, the OAP ratings might be considered more valid than simple %Correct measures (or at least as more encompassing), since they tend not to be biased by peripheral information, and they tend to reflect qualitative aspects of performance that are not tapped by a %Correct measure. That is, results suggest that work sample administrator global performance ratings are (about) as valid as has been presumed. However, we urge caution in generalizing the current findings to all work samples too readily for four reasons.

First, we know of only three other studies to bear on the issue of work sample validity (Borman & Hallam, 1991; Brugnoli et al., 1979; Hedge & Teachout, 1992), so although empirical evidence is encouraging, it is still very limited. Second, the present results stem from work sample test batteries that were developed using state of the technology precision. Every step in the work sample test battery development and administrator training followed from scientifically established principles in the job analytic, psychometric, and performance appraisal literatures. In this sense, the present research context may be as good as it gets, and our findings should not be generalized to other settings in which work sample development follows more ad hoc procedures.

Third, the work sample measurement process in the current study was actually a combination of the scoring schemes described earlier by F. D. Smith (1991), and most closely resembled the behavioral recording forms approach in which the recording of task step-level

performance information assists accurate OAP ratings. Consequently, our findings should not be readily generalized to situations in which only OAP ratings are obtained. Nevertheless, the present study's findings are the first to suggest that these ratings really are as valid as has been presumed.

Finally, our findings should not be generalized to other performance measurement situations that bear some (prhaps superficial) similarities to the work samples studied here. For example, many assessment center (AC) exercises bear resemblances to work samples, and post-exercise dimensional ratings (PEDRs) often closely resemble the OAPs reported in the present study. PEDRs typically are made using the global rating approach discussed by F. D. Smith (1991) in which summary judgments of (dimensional) performance are made following the completion of task (i.e., exercise) performance. However, AC exercises are usually much less structured (e.g., in terms of the specification of intermediate performance steps) than the work sample items investigated here, and we know of no research that has linked PEDRs to actual assessee behaviors as the OAPs were in the present study. We see this as a need for future, related, research.

Nevertheless, our findings seem to lend assurance to one of our "folk assumptions" regarding the type of criterion measures that are typically obtained from work samples: work sample administrator global task performance ratings appear to be (about) as valid as has been assumed. Good news!

References

Asher, J. T., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. Personnel Psychology, 27, 519-533.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol. 2) (pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Borman, W. C., & Hallam, G. L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-difference correlates. Journal of Applied Psychology, 76, 11-18.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp. 71-98). San Francisco: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10, 99-109.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of task performance and interpersonal factors on supervisor and peer performance ratings. Journal of Applied Psychology, 80, 168-177.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76, 863-872.

Brugnoli, G. A., Campion, J. E., & Basen, J. A. (1979). Racial bias in the use of work samples for personnel selection. Journal of Applied Psychology, 64, 119-123.

Burtch, L. D., Lipscomb, M. S., & Wissman, D. J. (1982). Aptitude requirements based on task difficulty: Methodology for evaluation. (AFHRL-TR-81-34) Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol. 1) (pp.687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. Personnel Psychology, 43, 313-333.

Campion, J. E. (1972). Work sampling for personnel selection. Journal of Applied Psychology, 56, 40-44.

Cooksey, R. W. (1996). Judgment analysis: Theory, methods, and applications. San Diego, CA: Academic Press.

Department of Defense (1984). Test manual for the armed services vocational aptitude battery. North Chicago, IL: United States Military Entrance Processing Command.

DuBois, C. L., Sackett, P. R., Zedeck, S., & Fogli, L. (1993). Further exploration of typical and maximum performance criteria: Definitional issues, prediction, and white-black differences. Journal of Applied Psychology, 78, 205-211.

Ford, J. K., Kraiger, K., & Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. Psychological Bulletin, 99, 330-337.

Hamner, C. W., Kim, J. S., Baid, L., & Bigoness, W. J. (1974). Race and sex as determinants ratings by potential employers in a simulated work sample task. Journal of Applied Psychology, 59, 705-711.

Hedge, J. W., Dickinson, T. L., & Bierstedt, S. A. (1988). The use of videotape technology to train administrators of walk-through performance testing. (AFHRL-TP-87-71). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., & Teachout, M. S. (1986). Job performance measurement: A systematic program of research and development. (AFHRL-TP-86-37). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Hedge, J. W., & Teachout, M. S. (1992). An interview approach to work sample criterion measurement. Journal of Applied Psychology, 77, 453-461.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. Psychological Bulletin, 57, 116-131.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. Research in Organizational Behavior, 5, 141-197.

Kavanagh, M. J., Borman, W. C., Hedge, J. W., & Gould, R. B. (1987). Job performance measurement in the military: A classification scheme, literature review, and directions for research. (AFHRL-TR-87-15). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Klimoski, R. , & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. Personnel Psychology, 40, 243-260.

Lance, C. E., Hedge, J. W., & Alley, W. E. (1989). Joint relations of task proficiency with ability, experience, and task difficulty: A cross-level, interactional study. Human Performance, 2, 249-272.

Lance, C. E., LaPointe, J. A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rater error. Journal of Applied Psychology, 79, 332-340.

Lance, C. E., Parisi, A. G., Bennett, W. Jr., Teachout, M. S., Harville, D. L., & Welles, M. L. (1998). Moderators of skill retention interval/performance decrement relationships in eight U.S. Air Force enlisted specialties. Human Performance, 11, 103-123.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. Journal of Applied Psychology, 77, 437-452.

Lance, C. E., Woehr, D. J., & Fisicaro, S. A. (1991). Cognitive categorization processes in performance evaluation: Confirmatory tests of two models. Journal of Organizational Behavior, 12, 1-20.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Laue, F. J., Hedge, J. W., Wall, M. L., Pederson, L. A., & Bentley, B. A. (1992). Job performance measurement system development process. (AL-TR-1992-0120). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Technical Training Research Division.

Martell, R. F., Guzzo, R. A., & Willis, C. E. (1995). A methodological and substantive note on the performance-cue effect in ratings of work-group behavior. Journal of Applied Psychology, 80, 191-195.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. Human Performance, 10, 71-83.

Motowidlo, S. J., & VanScotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. Journal of Applied Psychology, 79, 475-480.

Mumford, M. D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1987). Measuring occupational difficulty: A construct validation against training criteria. Journal of Applied Psychology, 72, 578-587.

Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. Journal of Applied Psychology, 74, 770-780.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. Journal of Applied Psychology, 79, 518-524.

Robertson, I. T., & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. Journal of Applied Psychology, 74, 402-410.

Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact, and applicant reaction. Journal of Occupational Psychology, 55, 171-183.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 88, 413-428.

Sackett, P. R., & DuBois, C. L. Z. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. Journal of Applied Psychology, 76, 873-877.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. Journal of Applied Psychology, 73, 482-486.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Smith, F. D. (1991). Work samples as measures of performance. In A. K. Wigdor & B. F. Green Jr. (Eds.), Performance assessment for the workplace (Vol. 2) (pp. 27-52). Washington, DC: National Academy Press.

Smith, M. (1994). A theory of the validity of predictors in selection. Journal of Occupational and Organizational Psychology, 67, 13-31.

Smith, P. C. (1976). Behavior, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp.745-775). Chicago: Rand McNally.

Teachout, M. S., & Pellum, M. W. (1991). Air Force research to link standards for enlistment to on-the-job performance. (AFHRL-TR-90-90). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.

Terpstra, D. E. (1996). The search for effective methods. HR Focus, 73(5), 16-17.

Thorndike, R. L. (1949). Personnel selection. New York: Wiley.

Thornton, G. C. (1992). Assessment centers in human resource management. Reading, MA: Addison Wesley.

Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. Academy of Management Journal, 28, 712-723.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. Journal of Applied Psychology, 81, 557-574.

Weeks, J. (1984). Occupational learning difficulty: A standard for determining the order of aptitude requirement minimums. (AFHRL-SR-84-26). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. Personnel Psychology, 35, 521-551.

Wigdor, A. K., & Green, B. F. Jr. (Eds.), (1991a). Performance assessment for the workplace (Vol. 1). Washington, DC: National Academy Press.

Wigdor, A. K., & Green, B. F. Jr. (Eds.), (1991b). Performance assessment for the workplace (Vol. 2). Washington, DC: National Academy Press.

Footnotes

[1]Data collection for the JPM project occurred in three sequential "waves." Data collection began with the Jet Engine Mechanic and Air Traffic Control Operator AFSs in the first wave, with data collection following for the remaining AFSs in subsequent waves. In the latter two waves, approximately 15% of the examinees' performance was evaluated using "shadow scoring," in which two test administrators independently observed and scored the examinee's step-level performance. Median interscorer reliabilities were $r = .97$ and $r = .93$ (Hedge & Teachout, 1992) for hands-on and interview work sample tasks (this distinction is described shortly), supporting the accuracy and objectivity of these step-level performance measures.

[2]Note that the work sample administrators did not themselves calculate the percentage of task steps completed correctly – they merely recorded whether each task step, individually, was completed correctly. Consequently, there was no direct mapping of some administrator-generated %Correct measure of task performance onto the 5-point global task performance rating scale.

[3]Items relating to examinee motivation were administered only in data collection waves two and three. Consequently, these data were unavailable for the Jet Engine Mechanic and Air Traffic Control samples.

[4]Descriptive statistics for each AFS separately are available from the first author.

[5]Data regarding statistically significant interaction effects are available from the first author.

[6]No shadow data were collected for the Jet Engine Mechanic AFS.

Table 1

Example Work Sample Task and Constituent Task Steps

Task: Installation of engine pressure ratio probes (Task #359).

Task steps:

1. Insert the pressure sensing probes into the turbine exhaust case.
2. Install the bolts and nuts into the turbine exhaust case bosses.
3. Connect the tube and manifold assemblies into the sensing probes.
4. Torque the probe nuts.
5. Torque the manifold and probe connection B nuts.
6. Install the safety device on the B nut.
7. Install the brackets and clips to the rear turbine exhaust case.

Table 2

Study Variables' Descriptive Statistics and Intercorrelations

			Variable									
Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. OAP	2.48	1.19	1.00									
2. %Correct	.67	.29	.77	1.00								
3. NTP	3.50	2.10	.33	.38	1.00							
4. TIME	6.65	6.97	-.18	-.18	-.18	1.00						
5. LTP	14.82	24.98	-.14	-.12	-.37	.11	1.00					
6. MOT	3.80	.66	.07	.06	.04	-.01@	-.03	1.00				
7. #STEPS	12.01	7.06	-.14	-.10	-.11	.52	.10	-.02@	1.00			
8. White	.80	.40	-.01@	-.02	-.01@	.05	-.01@	-.05	.03	1.00		
9. Black	.13	.34	.02	.02	.02	-.04	.01@	.03	-.03	-.80	1.00	
10. Sex	.85	.36	-.05	-.04	-.01@	.08	.04	.04	.06	.12	-.14	1.00
11. H/I	.55	.50	.09	.14	.03	.38	.01@	.00@	.12	.00@	.00@	.00@

Note. OAP = Overall Performance Rating; %Correct = Percentage of Task Steps Completed Correctly; NTP = Number of Times Performed; TIME = Time to Complete work sample task; LTP = Last Time Performed; MOT = Examinee Motivation; Black and White (= 1, versus Other racial groups = 0); Sex (1 = Male, 0 = Female); H/I = hands-on (=1) vs. interview (=0) administration mode.

@$p > .01$.

Table 3

Percentages of Statistically Significant Regression Weights for Predictors of OAP Ratings

Job	No. of Tasks	%Correct	NTP	TIME	LTP	MOT	Black	White	Sex
Avionics Communication	19	100.0 (0.0)	15.8 (0.0)	36.8 (10.5)	0.0 (0.0)	0.0 (0.0)	5.3 (0.0)	0.0 (5.3)	0.0 (0.0)
Air Traffic Control	14	100.0 (0.0)	0.0 (7.1)	14.3 (0.0)	0.0 (7.1)	N/A[a]	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
AGE Mechanic	31	100.0 (0.0)	6.5 (0.0)	0.0 (32.3)	3.2 (6.5)	0.0 (3.2)	0.0 (0.0)	0.0 (3.2)	3.2 (3.2)
Personnel	11	100.0 (0.0)	9.1 (0.0)	27.3 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (9.1)
Precision Measurement	21	100.0 (0.0)	4.8 (0.0)	52.4 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Aircrew Life Support	14	78.6 (0.0)	28.6 (0.0)	7.1 (7.1)	21.4 (0.0)	28.6 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Jet Engine Mechanic	8	100.0 (0.0)	0.0 (0.0)	0.0 (12.5)	0.0 (0.0)	N/A	0.0 (0.0)	0.0 (12.5)	0.0 (0.0)

Information Systems Radio Operator	16	100.0 (0.0)	6.3 (6.3)	0.0 (6.3)	6.3 (12.5)	0.0 (0.0)	6.3 (0.0)	0.0 (6.3)	0.0 (0.0)
Mean	134	97.8 (0.0)	9.0 (1.5)	17.9 (11.2)	3.7 (3.7)	3.0 (0.7)	1.5 (0.0)	0.0 (3.0)	0.7 (1.5)

Note. OAP = Overall Performance Rating; %Correct = Percentage of Task Steps Completed Correctly; NTP = Number of Times Performed; TIME = Time to Complete work sample task; LTP = Last Time Performed; MOT = Examinee Motivation; Black and White (= 1, versus Other racial groups = 0); Sex (1 = Male, 0 = Female). Numbers outside (inside) parentheses represent the percentages of regression coefficients that were statistically significant and in the hypothesized (opposite) direction.

[a]Examinee motivation was not assessed in the Air Traffic Control or Jet Engine Mechanic samples; see Footnote 2.

Table 4

Mean $\underline{R}^2$s and Standardized Regression Weights (βs)

	R^2	%Correct	NTP	TIME	LTP	MOT	Black	White	Sex
Mean	.54	.69	.04	-.03	.01	.02	-.01	-.01	.00
-1SD	.35	.48	-.04	-.17	-.07	-.04	-.09	-.10	-.06
+1SD	.68	.82	.12	.11	.09	.08	.07	.08	.06

Note. OAP = Overall Performance Rating; %Correct = Percentage of Task Steps Completed Correctly; NTP = Number of Times Performed; TIME = Time to Complete work sample task; LTP = Last Time Performed; MOT = Examinee Motivation; Black and White (= 1 versus Other racial groups = 0); Sex (1 = Male, 0 = Female).

Table 5

Prediction of OAP from %Correct and Other Factors Possibly Related to Work Sample Performance Ratings

Variable	β	t-ratio	R^2	F	ΔR^2	F
Step 1:						
%Correct (PC)	.759	142.86***	.577	20,410.05***	----	----
Step 2:						
MOT	.021	3.89***				
#STEPS	-.025	-4.01***				
LTP	-.034	-6.08***				
NTP	.032	5.41***				
TIME	-.022	-3.16**				
H/I	.006	n.s.				
Black	.018	2.01*				
White	.027	2.99**				
Sex	-.015	-2.80**	.583	2,092.72***	.006	24.92***
Step 3:						
PCxTIME	-.091	-13.96***				
PCx#STEPS	.093	14.55***				
PCxNTP	.038	6.08***				
PCxLTP	-.024	-4.16***				
H/IxTIME	-.061	-5.54***	.594	1,286.95***	.011	57.18***
Step 4:						
PCxH/IxTIME	-.070	-7.94***	.596	1,223.99***	.002	62.86***

Note. OAP = Overall Performance Rating; %Correct = Percentage of Task Steps Completed Correctly; MOT = Examinee Motivation; #STEPS = number of constituent task steps; LTP = Last Time Performed; NTP = Number of Times Performed; TIME = Time to Complete work sample task; H/I = hands-on (=1) vs. interview (=0) administration mode; Black and White (= 1, versus Other racial groups = 0); Sex (1 = Male, 0 = Female). * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 6

Prediction of OAP from %Correct-Shadow and Other Factors Possibly Related to Work Sample Performance Ratings

Variable	β	t-ratio	R^2	F	ΔR^2	F
Step 1:						
%Correct – Shadow (PCS)	.675	37.21***	.455	1,384.50***	----	----
Step 2:						
MOT	-.014	n.s.				
#STEPS	-.033	n.s.				
LTP	-.008	n.s.				
NTP	.030	n.s.				
TIME	-.052	-2.27*				
H/I	.031	n.s.				
Black	-.041	n.s.				
White	-.073	-2.44*				
Sex	.033	n.s.	.465	143.35***	.010	3.42**
Step 3:						
PCSxTIME	-.078	-3.64***				
PCSx#STEPS	.088	4.04***	.471	122.37***	.006	9.34***
Step 4:						
PCSxH/IxTIME	.022	n.s.	.471	112.99***	<.001	n.s.

Note. OAP = Overall Performance Rating; %Correct-Shadow = Percentage of Task Steps Completed Correctly – Shadow scores; MOT = Examinee Motivation; #STEPS = number of constituent task steps; LTP = Last Time Performed; NTP = Number of Times Performed; TIME = Time to Complete work sample task; H/I = hands-on (=1) vs. interview (=0) administration mode; Black and White (= 1, versus Other racial groups = 0); Sex (1 = Male, 0 = Female). * $p < .05$, ** $p < .01$, *** $p < .001$.

VALIDATION OF THE MULTIDIMENSIONAL WORK ETHIC PROFILE (MWEP) AS A SCREENING TOOL FOR AIR FORCE ENLISTED PERSONNEL

David J. Woehr
Associate Professor

Department of Psychology
Texas A&M University
College Station, Texas 77843-4235

Final Report for:
1998 Summer Research Extension Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

December 1998

VALIDATION OF THE MULTIDIMENSIONAL WORK ETHIC PROFILE (MWEP) AS A SCREENING TOOL FOR AIR FORCE ENLISTED PERSONNEL

David J. Woehr
Associate Professor

Department of Psychology
Texas A&M University
College Station, Texas 77843-4235

Abstract

The present study examines the psychometric properties of the Multidimensional Work Ethic Profile (MWEP) developed by Michael Miller and David Woehr (Woehr & Miller, 1997, Miller and Woehr, 1997) with Air Force enlisted personnel. The MWEP is a multidimensional measure of work ethic based on previous literature and research focusing on work ethic and job performance. Originally developed based on a sample of university students, the MWEP has demonstrated good psychometric characteristics including reliability and validity. The MWEP has been suggested as a potentially valuable screening tool with Air Force enlisted personnel. The purpose of the present study was to provide a preliminary evaluation of the measure among Air Force enlisted personnel. Results indicate that the measure does demonstrate similar psychometric characteristics among Air Force enlisted personnel as with the original developmental sample. The MWEP provides reliable and valid measures of multiple dimensions underlying the work ethic construct. These results indicate that the MWEP may be a useful screening tool for Air Force Personnel.

VALIDATION OF THE MULTIDIMENSIONAL WORK ETHIC PROFILE (MWEP) AS A SCREENING TOOL FOR AIR FORCE ENLISTED PERSONNEL

David J. Woehr
and
Michael J. Miller
Texas A&M University

Introduction

History and Definition of Work Ethic

The term "work ethic" was coined centuries ago by post-Reformation intellectuals who opposed the practice of social welfare and professed the importance of individualism (Byrne, 1990). They espoused the belief that human beings must assume full responsibility for their lot in life and the poor were no exception. As such, hard work was viewed as a panacea and through it, one could improve his or her condition in life. Implicit in this assumption was the belief that the poor simply needed to help themselves through diligent labor and all life's ills would vanish. Such were the harsh origins of the construct.

Modern formulations of the work ethic construct stem from the work of the German scholar Max Weber. It was in 1904 and 1905 that Weber wrote a two-part essay entitled "The Protestant Ethic and the Spirit of Capitalism". In this essay Weber advanced the thesis that the introduction and rapid expansion of capitalism and the resulting industrialization in Western Europe and North America was <u>in part</u> the result of the Puritan value of asceticism (i.e., scrupulous use of time, strict self-denial of luxury, worldly pleasure, ease, and so on to achieve personal discipline) and the belief in a calling from God (Byrne, 1990; Charlton, Mallinson, & Oakeshott, 1986; Fine, 1983; Furnham, 1990a; Green, 1968; Lehmann, 1993; Maccoby, 1983; Nord, Brief, Atieh, & Doherty, 1988; Poggi, 1983). It was the practice of asceticism that Weber believed produced the celebrated 'work ethic'--the complete and relentless devotion to one's economic role on earth (Lessnoff, 1994). An individual's economic role was prescribed by the belief in a calling (Gilbert, 1977). The manifestation of occupational rewards through success in one's calling came to be revered as a sign of being one of the elect (i.e., chosen by God to receive salvation). Thus, economic activity was a vehicle toward economic success and economic success was a sign of

salvation.

Weber maintained that other Protestant faiths (e.g., Calvinism, Methodism, Pietism, and Baptists) shared common theological underpinnings in terms of being proponents of asceticism and the spirit of capitalism (Bouma, 1973; Nelson, 1973); thus the term "Protestant Work Ethic" (PWE). However, the premise that work ethic is a religiously oriented concept was contested then and since. In fact, researchers have found little relationship between religious orientation and endorsement of the work ethic (Giorgi & Marsh, 1990; Ray, 1982). Ray (1982) concluded that all religious orientations currently share the attributes associated with the work ethic to the same degree. He states that the Protestant ethic, "...is certainly not yet dead; it is just no longer Protestant" (p. 135). This is consistent with Pascarella's (1984) contention that all major religions have espoused the importance of work. Thus, it appears that what was originally conceived as a religious construct is now likely secular and is best viewed as general work ethic and not the PWE.

Since work ethic is not a surrogate for religious orientation the question becomes, What is it? Current conceptualizations tend to view work ethic as an attitudinal construct pertaining to work oriented values. An individual espousing a high work ethic would place great value on: hard work, autonomy, fairness, wise and efficient use of time, delay of gratification, and the intrinsic value of work (Cherrington, 1980; Dubin, 1963; Furnham, 1984; Ho & Lloyd, 1984; Weber, 1958; Wollack, Goodale, Wijting, & Smith, 1971). Therefore, work ethic seems to be made up of multiple components. These components appear to include: industriousness, asceticism, self-reliance, morality, delay of gratification, and the centrality of work. In the absence of a firmly accepted conceptual and operational definition it is posited that work ethic is a construct that reflects a constellation of attitudes and beliefs pertaining to work oriented behavior. Characteristics of "work ethic" are that it: (a) is multidimensional; (b) pertains to work and work related activity in general, not specific to any particular job (yet may generalize to domains other than work - school, hobbies, etc.); (c) is learned (not dispositional); (d) refers to attitudes and beliefs (not necessarily behavior); (e) is intendend as a motivational construct (should be reflected in behavior); and (e) is secular, not necessarily tied to any one set of religious beliefs.

Relevance of Work Ethic to the Air Force

As previously defined, individual differences in work ethic should reflect differences among individuals in

terms of their attitudes and beliefs with respect to the value of work and work-related behavior. An important consideration for industrial psychology is the relationship between these attitudes and beliefs and actual work behavior. While industrial psychologists interested in the work ethic have typically explored its relationship with other attitudinal variables such as job satisfaction (e.g., Aldag & Brief, 1975; Blood, 1969; Stone, 1975, 1976; Wanous, 1974), job involvement (e.g., Blau, 1987; Randall & Cote, 1991; Saal, 1978), and organizational commitment (e.g., Kidron, 1978; Morrow & McElroy, 1987), there have been relatively few studies (e.g., Khaleque, 1992; Orpen, 1986), focusing on the relationship of work ethic with actual job performance. A possible reason for this is the lack of distinction between task and contextual aspects of job performance.

Recently several models of job performance have been proposed which attempt to describe a set of underlying dimensions that are representative of performance in all jobs (Borman & Motowidlo, 1993; Campbell, 1990; Campbell, McCloy, Oppler, & Sager, 1993). For example, Campbell (1990) argues that all jobs are made up of eight factors, including: job-specific task proficiency, non-job-specific task proficiency, written and oral communication, demonstrating effort, maintaining personal discipline, facilitating team and peer performance, supervision and leadership, and management and administration. Campbell's formulation distinguishes between behaviors that contribute to organizational effectiveness through their focus on task proficiency and those behaviors that help the organization in other ways (Motowidlo & Van Scotter, 1994). Task proficiency behaviors are formally prescribed by the organization whereas other behaviors, though not formally a part of the job, are still very valuable for organizational effectiveness (Borman & Motowidlo, 1993).

Borman and Motowidlo (1993) place performance behaviors not prescribed by the organization under the rubric of contextual activities. Examples include:

(1) Volunteering to carry out task activities that are not formally a part of the job.
(2) Persisting with extra enthusiasm or effort when necessary to complete own task activities successfully.
(3) Helping and cooperating with others.
(4) Following organizational rules and procedures even when personally inconvenient.
(5) Endorsing, supporting, and defending organizational objectives. (p. 73)

Using a sample comprising Air Force mechanics, Motowidlo and Van Scotter (1994) demonstrated that supervisors consider task performance and contextual performance separately when providing performance ratings.

It is the contextual component of job performance in which work ethic may offer substantial predictive utility. Specifically, it may be possible to predict with a measure of work ethic the extent to which an individual would engage in contextual performance of value to the unit. Further, the work ethic may demonstrate a relationship with technical school training success, job performance, and tenure in the Air Force.

Measurement of Work Ethic

Of paramount concern for research focusing on the understanding of the work ethic construct as well as the relationship between work ethic and work behavior is the ability to accurately measure the construct. There are at least seven work ethic measures in existence which purport to provide reliable and valid measures of this construct. However, there are a number of problems with these measures. First and foremost, they focus on the measurement of a single construct by providing a global "work ethic" score. This is a considerable shortcoming as, since its inception, Weber believed the work ethic to be a multidimensional construct; a position that has subsequently been supported by numerous researchers (Bouma, 1973; Cherrington, 1980; Furnham, 1984; Oates, 1971).

From a psychometric as well as a conceptual perspective, the lack of focus on the multidimensional nature of the work ethic is troubling. The use of a single overall score could potentially cause the loss of information with regards to the different components of work ethic as well as their relationships with other constructs (Carver, 1989; McHoskey, 1994). Further, the use of a single score in studies using different instruments to measure the work ethic may at least partially explain the equivocal results often found in the literature (Furnham, 1984). That is, one cannot be sure if the conflicting results are due to a lack of robustness in the studies, the scales measuring different components of the work ethic, or deficiencies in terms of construct relevance and psychometric properties (Furnham, 1990b).

A second concern is that the various measures appear to tap different components of the work ethic and not the construct in its entirety. This has often led to poor intercorrelations among measures. For example, Furnham (1990b) administered seven measures of the work ethic to 1,021 participants and found that the correlations between the various measures ranged from 0.19 - 0.66 with a mean r of 0.36. One would expect the values to be much higher if the scales were indeed measuring the same thing.

Finally, another potential problem with existing work ethic measures is that these measures are relatively dated. The mean time since publication for the previous measures is 23 years. The age of the measures poses the problem of many dated items. For example, some of the items contain sex-biased language such as: "Hard work makes a man a better person", "The man who can approach an unpleasant task with enthusiasm is the man who gets ahead", and "To be superior a man must stand alone".

Factor analytic investigations of the various measures have found the existence of several identifiable factors (Furnham, 1990b; Heaven, 1989; Tang, 1993; Mirels & Garrett, 1971; McHoskey, 1994). For example, McHoskey (1994) factor analyzed Mirels and Garrett's Protestant Ethic scale. His analysis yielded a 4-factor solution which he labeled, "success", "asceticism", "hard-work", and "anti-leisure". However, McHoskey was quick to point out that though this scale was multidimensional, other important aspects of the PWE were absent. Specifically, it in no way measured an individual's attitudes toward morality, self-reliance, or delay of gratification. This lack of comprehensiveness in measuring the work ethic has been levied against other scales as well and limits their utility (Furnham, 1984, 1990a, b; McHoskey, 1994).

In an effort to ameliorate the shortcomings in previous attempts to measure the work ethic, Woehr and Miller (1997) and Miller and Woehr (1997) developed the Multidimensional Work Ethic Profile (MWEP). The goal in the development of such a measure was to build on and extend previous measures in an attempt to capture the multidimensionality of the construct. The MWEP is a 65-item measure assessing 7 dimensions related to the work ethic construct. These dimensions are: *"Delay of Gratification"*, *"Hard Work"*, *"Morality/Ethics"*, *"Self-Reliance"*, *"Leisure"*, *"Wasted Time"*, and *"Centrality of Work"*. Complete definitions of these dimensions are provided in table 1.

Originally developed based on a sample of university students, the MWEP has demonstrated good psychometric characteristics including reliability and validity. Specifically, Miller and Woehr (1997) report 3 - 4 week test-retest reliabilities of 0.83 - 0.95 and internal consistency coefficient alphas of 0.78 - 0.89 for the dimensions of work ethic. With regards to construct-related validity the MWEP demonstrated discriminant relationships with personality, cognitive ability, and manifest needs. Lastly, the criterion-related validity of the MWEP was evaluated by relating it to academic effort indices pertinent to the university student sample. The

MWEP was shown to be significantly related to hours studying per week (0.21), hours watching TV per week (0.36), hours in extracurricular activities per week (0.26), and classes missed (0.30).

Table 1.

Dimension definitions for the 7 work ethic dimensions assessed by the MWEP.

Dimension:	**Definition:**
Centrality of Work	Belief in the virtues of hard work.
Delay of Gratification	Striving for independence in one's daily work.
Hard Work	Pro-leisure attitudes and beliefs in activities that serve a rejuvenating function.
Leisure	Belief in work for work's sake and the importance of work.
Morality/Ethics	Believing in a just and moral existence.
Self-Reliance	Orientation toward the future; the postponement of rewards.
Wasted Time	Attitudes and beliefs reflecting active and productive use of time.

Present Study

Given the previous evaluations of the MWEP and the potential for use as a screening measure among Air Force enlisted personnel, the objective of this study was to empirically determine the extent to which the psychometric properties of the MWEP that have been found with a university student sample would generalize to Air Force enlisted personnel. Measurement stability across the samples would allow for greater confidence with regards to measurement equivalence and provide an initial indication of the viability of the MWEP for use in the Air Force.

As noted, the primary objective of the present study was to compare the psychometric characteristics of the MWEP with Air Force personnel relative to the original student development sample. This comparison focused on: (1) the mean score levels on each dimension, (2) score variability for each dimension, (3) the reliability for each dimension, and (4) the overall pattern of correlations among dimensions. If the MWEP functions similarly across the two samples no differences in dimension variability, dimension reliability, or the overall pattern of

correlations among dimensions should be found. However, differences in mean levels on each dimension are likely given the actual differences across the two samples. That is, the student sample represents 18 to 22-year-old college students. Alternately, the Air Force sample represents an 18 to 22-year-old non-college bound sample. It is likely that actual differences in work ethic attitudes and beliefs exist across the two groups. Such differences would be reflected in mean dimension score differences.

Method

University Participants.

The university student sample comprised 598 participants (52% female and 48% male). Subject participation was voluntary and subjects received partial course credit for taking part in the study. Mean age of the participants was 19.2 and ranged from 17 to 27.

Air Force Participants.

Participants in the present study were 741 Air Force enlisted personnel that participated in the study during Basic Military Training (BMT). The participants were 60% male and 40% female. Further, 70% were White, 20% Black, 6% Hispanic, 3% Asian, and 1% Other. Mean age of the participants was 18.76 and ranged from 18 to 28.

Multidimensional Work Ethic Profile (MWEP) Measure.

The MWEP was originally developed as a 65 item paper-and-pencil measure. The measure requires responses to items on a 5 point Likert-type scale ranging from 1(strongly disagree) to 5 (strongly agree). In order to facilitate data collection in the present study the MWEP was included as part of a computer-administered battery of questionnaires. Thus a computer administered version of the MWEP was developed. Although computer-administered this version was highly similar to the paper-and-pencil version. Both items and response options were displayed in the same manner in both forms. Participants were asked to respond to each of the items via the numbers on the computer keyboard.

Procedure.

The MWEP was administered as part of an extensive battery of computer-administered questionnaires completed in a single 4 hour session during the first week of BMT. Subjects were seated at individual computer

terminals and given the measures. Administration of the measures was counterbalanced across experimental sessions.

Results

Comparison of the MWEP in the two samples focused on: (1) the mean score levels on each dimension, (2) score variability for each dimension, (3) the reliability for each dimension, and (4) the overall pattern of correlations among dimensions. Mean scores for each of the 7 work ethic dimensions for both the Air Force and student samples are presented in Table 2.

Table 2.

Means and standard deviations for the 7 work ethic dimensions for both the Student and Air Force Samples.

	Student Sample N = 598		**Air Force Sample** N = 741		
Dimension:	**Mean**	**SD**	**Mean**	**SD**	t
Centrality of Work	24.37	6.04	20.33	5.77	12.47*
Delay of Gratification	24.29	6.43	19.42	5.76	14.42*
Hard Work	22.10	5.87	16.41	5.23	18.76*
Leisure	31.32	5.86	27.91	5.75	10.69*
Morality/Ethics	16.10	4.47	13.49	3.22	11.98*
Self-Reliance	26.15	6.84	24.48	7.13	4.35*
Wasted Time	24.98	5.89	20.08	5.37	15.88*
Total Score	169.31	25.43	142.12	24.70	19.70*

*$p < .01$.

Tests for differences between the mean scores for each dimension are also presented in Table 2. These results indicate significant mean differences for all dimensions. Further, means are higher for the student sample than for the Air Force sample for all dimensions.

Table 3 provides the results of a comparison of the variance of each dimension across samples. These results indicate no significant differences (at the $p \leq .01$ level)for any of the dimensions across samples except

"Morality/Ethics" and *"Delay of Gratification"*. For both the morality/ethics and delay of gratification dimensions there is significantly less variability in scores for the Air Force sample than for the student sample.

Table 3.

Test for equality of variances across student and Air Force Samples.

			Levine's Test for Equality of Variances	
Dimension	**Student Sample Variance**	**Air Force Sample Variance**	**F**	**p**
Centrality of Work	36.47	36.34	.927	.336
Delay of Gratification	41.38	33.22	6.84	.009
Hard Work	34.41	27.38	3.643	.057
Leisure	34.30	33.09	.338	.561
Morality/ethics	19.96	10.40	52.751	.000
Self-Reliance	46.82	50.79	2.26	.133
Wasted Time	34.66	28.86	3.956	.047
Total Score	646.79	609.96	.003	.953

Dimension reliabilities (coefficient α) for both samples are presented in table 4. Examination of these results indicate no differences in dimension reliabilities across samples except for the *"Morality/Ethics"* dimension. Specifically, all dimension reliabilities are within .03 of each other across samples except for the *"Morality/Ethics"* dimension for which the reliability is substantially lower in the Air Force sample.

Finally, the dimension intercorrelations for both the Air Force and student samples are presented in table 5. In order to assess the extent to which the dimension intercorrelations differed across samples, we used LISREL 8.14 (Joreskog & Sorbom, 1993) to provide an overall test of the equivalence of the 2 correlation matrices. Specifically, we tested a model in which correlations among the 7 work ethic dimensions were set equal to the

student sample based correlations and the correlations for the Air Force sample were constrained to be equal to those from the student sample. Using this approach, the overall model fit indices derived from the LISREL analyses provide an indication of the overall equality of the correlations across samples. Results of this analysis are provided in Table 6 and indicate that the two sets of correlations are generally equivalent.

Table 4.

Examination of reliabilities across student and Air Force Samples.

Dimension	Student Sample Reliability	Air Force Sample Reliability
Centrality of Work	.84	.84
Delay of Gratification	.79	.77
Hard Work	.85	.86
Leisure	.87	.86
Morality/ethics	.78	.57
Self-Reliance	.89	.87
Wasted Time	.79	.76

Discussion

The present study presents an examination of the psychometric properties of the Multidimensional Work Ethic Profile (MWEP) developed by Michael Miller and David Woehr (Woehr & Miller, 1997, Miller and Woehr, 1997). The MWEP is a 65 item measure of work ethic based on previous research and literature focusing on work ethic and job performance. An important characteristic of the MWEP is that it assess 7 conceptually and empirically distinct facets of the work ethic construct. Originally developed based on a sample of university students, the MWEP has demonstrated good psychometric characteristics including reliability and convergent and discriminate validity. Further, the MWEP has been suggested as a potentially valuable screening tool with Air Force enlisted personnel. The purpose of the present study was to provide a preliminary evaluation of the measure

among Air Force enlisted personnel. Results indicate that the measure does in fact demonstrate highly similar psychometric characteristics among Air Force enlisted personnel as with the original developmental sample. The MWEP provides reliable and valid measures of multiple dimensions underlying the work ethic construct. These results indicate that the MWEP may be a useful screening tool for Air Force Personnel.

Table 5.

Work ethic dimension intercorrelations for the student and Air Force samples.

	Student Sample						
Dimensions:	1	2	3	4	5	6	7
1. Centrality of Work	1.0						
2. Delay of Gratification	.38	1.0					
3. Hard Work	.33	.33	1.0				
4. Leisure	-.47	-.12	-.08	1.0			
5. Morality/Ethics	.17	.25	.22	.08	1.0		
6. Self-Reliance	.20	.21	.38	.10	.13	1.0	
7. Wasted Time	.56	.40	.38	-.28	.21	.32	1.0
	Air Force Sample						
1. Centrality of Work	1.0						
2. Delay of Gratification	.47	1.0					
3. Hard Work	.50	.56	1.0				
4. Leisure	.42	.28	.26	1.0			
5. Morality/Ethics	.34	.38	.43	.20	1.0		
6. Self-Reliance	.16	.16	.16	.07	.11	1.0	
7. Wasted Time	.60	.51	.59	.29	.38	.20	1.0

Student Sample $N = 598$. All correlations are significant ($p < .01$).

Air Force Sample $N = 741$. All correlations greater than .11 are significant ($p < .01$).

Table 6.

Goodness of fit indices for the test of intercorrelation equivalence.

χ^2	df	χ^2/df	RMSEA	GFI	NFI	CFI	RFI
88.82	35	2.53	.05	.98	.96	.98	.95

Specifically, results of the present study found no differences across samples for the dimension variances, reliabilities, and intercorrelations across dimensions. One exception to these findings was for the *"Morality/Ethics"* dimension. For this dimension the results indicated significantly less variance as well as substantially lower reliability with the Air Force sample relative to the student sample. One possible explanation for this finding may lie in differences in the work settings of the two samples. That is, the student sample was assessed in a non-job setting while the Air Force sample was assessed in an actual job setting. It is likely that the items comprising the *"Morality/Ethics"* dimension are fairly transparent and actual job incumbents may not respond as truthfully as non incumbents. This would explain the restricted variance found in the Air Force sample. This reduced variance would in turn result in a lower reliability estimate. Counter to this explanation, however, was our finding that the mean response for the *"Morality/Ethics"* dimension was actually significantly lower in the Air Force sample relative to the student sample. If the items were relatively transparent and the incumbent sample was simply responding in a more socially desirable manner then one would expect a higher mean score. It is difficult at this point to determine the exact reasons for the differences found across samples for this dimensions. The lack of differences across the other, more work-related, dimensions however is encouraging.

The results of the present study do indicate significant mean score differences for all of the 7 dimensions across samples. These differences are not unexpected and do not call into question the measurement equivalence of the MWEP in either sample. Rather these differences are to a certain extent consistent with expected differences between the two samples. The student sample represents young adults attending college. Alternately, the Air Force sample represents young adults not attending college but directly entering the work force. Thus differences in work ethic scores most likely reflect actual differences between samples.

Conclusion

The prediction of job performance is one of the benchmarks of industrial psychology. Though the field has relied primarily on cognitive ability measures to predict performance, it has also pursued the use of alternative predictors (Arvey & Sackett, 1993). One of the most prevalent alternative predictors has been personality variables (Adler, 1996; Barrick & Mount, 1991; Goffin, Rothstein, & Johnston, 1996; Hogan, Hogan, & Roberts, 1996; Horman & Maschke, 1996; Tett, Jackson, & Rothstein, 1991). Though measures of personality have not resulted in adverse impact, many researchers have found a low relationship with actual criterion measures of job performance (Ones, Mount, Barrick, & Hunter, 1994). Another potential problem is that personality variables may not function in a linear fashion. Attitudinal variables such as work ethic may bridge the gap between cognitive ability and personality variables.

The present study demonstrates that one such attitudinal measure, the MWEP, a multidimensional measure of work demonstrates good psychometric characteristics in two diverse samples. This suggests that the MWEP is a potentially valuable pragmatic measure for either sample. Certainly, the next step is to examine the predictive utility of the MWEP in an Air Force context. An avenue of research for the future would be an examination of the relationship of the work ethic to technical school training success, job performance, and tenure in the Air Force. This could be achieved through the administration of the MWEP to enlisted personnel while in BMT and following up on their respective progress in the Air Force. The criteria in this example might be technical school final grades, performance evaluations while at the duty station, and fulfilment of enlistment tour requirements. Such a criterion-related validity study is currently in progress.

References

Aldag, R. J. & Brief, A. P. (1975). Some correlates of work values. Journal of Applied Psychology, 60, 757-760.

Blau, G. J. (1987). Using a person-environment fit model to predict job involvement and organizational commitment. Journal of Vocational Behavior, 30, 240-257.

Blood, M. R. (1969). Work values and job satisfaction. Journal of Applied Psychology, 53, 456-459.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp. 71-98). San Francisco: Jossey-Bass.

Bouma, G. D. (1973). Beyond Lenski: A critical review of recent "Protestant Ethic" research. Journal for the Scientific Study of Religion, 12, 141-155.

Byrne, E. F. (1990). Work, inc.: A philosophical inquiry. Philadelphia: Temple University Press.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol 1, 2nd ed., pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco: Jossey-Bass.

Carver, C. S. (1989). How should multifaceted personality constructs be tested? Journal of Personality and Social Psychology, 56, 577-585.

Charlton, W., Mallinson, T., & Oakeshott, R. (1986). The Christian response to industrial capitalism. London: Sheed & Ward.

Cherrington, D. J. (1980). The work ethic: Working values and values that work. New York: AMACOM.

Dubin, R. (1963). Industrial worker's worlds: A study of the 'central life interests' of industrial workers. In E. O. Smigel (Ed.), Work and Leisure (pp. 53-72). New Haven: College and University Press.

Fine, R. (1983). The Protestant ethic and the analytic ideal. Political Psychology, 4, 245-264.

Furnham, A. (1984). The Protestant work ethic: A review of the psychological literature. European Journal of Social Psychology, 14, 87-104.

Furnham, A. (1990a). The Protestant work ethic: The psychology of work-related beliefs and behaviours. London: Routledge.

Furnham, A. (1990b). A content, correlational, and factor analytic study of seven questionnaire measures of the Protestant work ethic. Human Relations, 43, 383-399.

Gilbert, J. B. (1977). Work without salvation: America's intellectuals and industrial alienation, 1880-1910. Baltimore: Johns Hopkins University Press.

Giorgi, L., & Marsh, C. (1990). The Protestant work ethic as a cultural phenomenon. European Journal of Social Psychology, 20, 499-517.

Green, A. W. (1968). Sociology: An analysis of life in modern society. New York: McGraw-Hill.

Ho, R., & Lloyd, J. I. (1984). Development of an Australian work ethic scale. Australian Psychologist, 19, 321-332.

Heaven, P. C. L. (1989). Structure and personality correlates of the Protestant work ethic among women. Personality and Individual Differences, 10, 101-104.

Joreskog, K. G., Sorbom, D. (1993). LISREL-8: Structural equation modeling with the SIMPLIS command language. Chicago, IL: Scientific Software International Inc.

Khaleque, A. (1992). Work values, attitudes and performance of industrial workers in Bangladesh. Social Indicators Research, 27, 187-195.

Kidron, A. (1978). Work values and organizational commitment. Academy of Management Journal, 21, 239-247.

Lehmann, H. (1993). The rise of capitalism: Weber versus Sombart. In H. Lehmann & G. Roth (Eds.), Weber's Protestant ethic: Origins, Evidence, Contexts (pp. 195-208). Washington, D. C.: Cambridge University Press.

Lessnoff, M. H. (1994). The spirit of capitalism and the Protestant ethic: An Enquiry into the Weber thesis. Brookfield, VT: E. Elgar.

Maccoby, M. (1983). The managerial work ethic in America. In J. Barbash, R. J. Lapman, S. A. Levitan, & G. Tyler (Eds.), The work ethic--A critical analysis (pp. 183-196). Madison, WI: Industrial Relations Research Association.

McHoskey, J. W. (1994). Factor structure of the Protestant work ethic scale. Personality and Individual Differences, 17, 49-52.

Mirels, H. L., & Garrett, J. B. (1971). The Protestant ethic as a personality variable. Journal of Consulting and Clinical Psychology, 36, 40-44.

Morrow, P. C., & McElroy, J. C. (1987). Work commitment and job satisfaction over three career stages. Journal of Vocational Behavior, 30, 330-346.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. Journal of Applied Psychology, 79, 475-480.

Nelson, B. (1973). WeberÕs Protestant ethic: Its origins, wanderings, and foreseeable future. In C. Glock & P. Hammond (Eds.), Beyond the classics (pp. 71-130). New York: Harper & Row.

Nord, W. R., Brief, A. P., Atieh, J. M., & Doherty, E. M. (1988). Work values and the conduct of organizational behavior. Research in Organizational Behavior, 10, 1-42.

Poggi, G. (1983). Calvinism and the capitalist spirit: Max Weber's Protestant ethic. Amherst: University of Massachusetts Press.

Miller, M. J., & Woehr, D. J. (1997). Work ethic: The development and evaluation of a multidimensional measure. Manuscript submitted for publication.

Oates, W. (1971). Confessions of a workaholic: The facts about work addiction. New York: World Publishing Company.

Pascarella, P. (1984). The new achievers: Creating a modern work ethic. New York: Free Press.

Randall, D. M. & Cote, J. A. (1991). Interrelationships of work commitment constructs. Work and Occupations, 18, 194-211.

Ray, J. J. (1982). The Protestant ethic in Australia. Journal of Social Psychology, 116, 127-138.

Saal, F. E. (1978). Job involvement: A multivariate approach. Journal of Applied Psychology, 63, 53-61.

Stone, E. F. (1975). Job scope, job satisfaction, and the Protestant ethic: A study of enlisted men in the U.S. Navy. Journal of Vocational Behavior, 7, 215-234.

Stone, E. F. (1976). The moderating effect of work-related values on the job scope-job satisfaction relationship. Organizational Behavior and Human Performance, 15, 147-167.

Tang, T. L. -P. (1993). A factor analytic study of the Protestant work ethic. The Journal of Social Psychology, 133, 109-111.

Wanous, J. P. (1974). Individual differences and reactions to job characteristics. Journal of Applied Psychology, 59, 616-622.

Weber, M. (1958). The Protestant Ethic and the spirit of capitalism (T. Parsons, Trans.). New York: Charles Scribner's Sons.

Woehr, D. J. & Miller, M. J. (1997, April). The meaning and measurement of work ethic. Paper presented at the annual Southeastern Industrial/Organizational Psychology Meeting, Atlanta, GA.

Wollack, S., Goodale, J. G., Wijting, J. P., & Smith, P.C. (1971). Development of the Survey of Work Values. Journal of Applied Psychology, 55, 331-338.